GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:26:23 ; Search time 3097 Seconds
(without alignments)
1954.997 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980054/runat_13022004_163839_26298/app_query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054_@CGN_1_1_3508_@runat_13022004_163839_26298 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		772	100.0	447	6	AX057345	AX057345 Sequence
2		764	99.0	447	6	AX240813	AX240813 Sequence
3		764	99.0	447	6	AX488762	AX488762 Sequence
4		562	72.8	326	6	AX057353	AX057353 Sequence
5	c	223	28.9	1104	8	SCYNL260C	Z71536 S.cerevisia
6		223	28.9	2961	8	YSCATX1HF	L35270 Saccharomyc
7	c	223	28.9	33016	8	SCCHXIVL	X96722 S.cerevisia
8		132	17.1	25000	8	SPCC191	AL049644 S.pombe c
9		130	16.8	105692	2	AP003953	AP003953 Oryza sat
10		130	16.8	144741	8	AP004674	AP004674 Oryza sat
11		122.5	15.9	96283	8	NC20H10	BX294024 Neurospor
12		111	14.4	1904	8	AF360255	AF360255 Arabidops
13		111	14.4	129667	8	AC006234	AC006234 Arabidops
14	c	108	14.0	58350	2	AC137683	AC137683 Homo sapi
15		107.5	13.9	30651	1	AE001581	AE001581 Borrelia
16	c	107	13.9	65	6	AX485584	AX485584 Sequence
17		104	13.5	29838	1	AE001578	AE001578 Borrelia
18	c	101.5	13.1	257109	3	AC115577	AC115577 Dictyoste
19		100.5	13.0	577	6	BD108716	BD108716 EST and e
20	c	100	13.0	76351	2	AC124635	AC124635 Mus muscu
21	c	100	13.0	171586	2	AC138292	AC138292 Mus muscu
22		99.5	12.9	445	6	BD025717	BD025717 Sequence
23		99.5	12.9	898	9	BC022043	BC022043 Homo sapi
24		99.5	12.9	901	9	AF226046	AF226046 Homo sapi
25		97.5	12.6	12190	1	AE001131	AE001131 Borrelia
26		97	12.6	21170	1	AF169008	AF169008 Borrelia
27	c	96	12.4	348600	1	AB063521	AB063521 Wiggleswo
28		95.5	12.4	2308	3	BT003595	BT003595 Drosophil
29	c	95.5	12.4	64195	2	AC015208	AC015208 Drosophil
30		95.5	12.4	184621	3	AC007440	AC007440 Drosophil
31		95.5	12.4	292919	3	AE003823	AE003823 Drosophil
32		95	12.3	2796	9	BC047097	BC047097 Homo sapi
33		95	12.3	6269	9	AB040962	AB040962 Homo sapi
34	c	95	12.3	138751	10	AC128666	AC128666 Mus muscu
35		95	12.3	182107	2	AC116694	AC116694 Mus muscu
36		94.5	12.2	30885	1	AE001580	AE001580 Borrelia
37		94	12.2	4239	1	BBU62901	U62901 Borrelia bu
38	c	94	12.2	10479	1	BBAJ3222	AJ003222 Borrelia
39		94	12.2	10845	1	AE001129	AE001129 Borrelia
40		93	12.0	207343	9	AC104389	AC104389 Homo sapi
41	c	93	12.0	299180	1	AP005080	AP005080 Vibrio pa
42	c	92	11.9	10185	1	AE004286	AE004286 Vibrio ch
43		92	11.9	73472	2	AC084770	AC084770 Homo sapi
44	c	92	11.9	201414	2	AC109398	AC109398 Rattus no
45	c	92	11.9	220718	2	AC130762	AC130762 Rattus no

ALIGNMENTS

RESULT 1

```
AX057345
LOCUS       AX057345                447 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 11 from Patent WO0075305.
ACCESSION   AX057345
VERSION     AX057345.1  GI:12310086
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Lalanne,J.L. and Rocher,C.
  TITLE     Novel candida albicans genes and proteins coded by said genes
  JOURNAL   Patent: WO 0075305-A 11 14-DEC-2000;
            HOECHST MARION ROUSSEL (FR)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
     CDS             1. .447
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=12
                     /protein_id="CAC22438.1"
                     /db_xref="GI:12310087"
                     /translation="MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTG
                     FQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEK
                     NIKKARNKLRVIASITKETWKIDSLDNLVKEVGGTLQVSENPDDMW"
BASE COUNT      192 a     38 c     82 g    135 t
ORIGIN
```

Alignment Scores:

Pred. No.:	7.61e-65	Length:	447
Score:	772.00	Matches:	148
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-980-054A-12 (1-148) x AX057345 (1-447)

```
Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        141 SerGluAsnProAspAspMetTrp 148
              |||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGG 444
```

```
RESULT 2
AX240813
LOCUS       AX240813                447 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 51 from Patent WO0160975.
ACCESSION   AX240813
VERSION     AX240813.1  GI:15797749
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 0160975-A 51 23-AUG-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      193 a     38 c     80 g    136 t
ORIGIN
```

Alignment Scores:

Pred. No.:	4.37e-64	Length:	447
Score:	764.00	Matches:	146
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	98.65%	Mismatches:	0
Query Match:	98.96%	Indels:	0
DB:	6	Gaps:	0

US-09-980-054A-12 (1-148) x AX240813 (1-447)

```
Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              ||||||:::|||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        141 SerGluAsnProAspAspMetTrp 148
              |||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGG 444
```

```
RESULT 3
AX488762
LOCUS       AX488762                447 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6062 from Patent WO02053728.
ACCESSION   AX488762
VERSION     AX488762.1  GI:22322774
```

```
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6062 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      193 a      38 c      80 g     136 t
ORIGIN

Alignment Scores:
Pred. No.:                4.37e-64        Length:         447
Score:                    764.00          Matches:        146
Percent Similarity:       100.00%         Conservative:   2
Best Local Similarity:    98.65%          Mismatches:     0
Query Match:              98.96%          Indels:         0
DB:                       6               Gaps:           0

US-09-980-054A-12 (1-148) x AX488762 (1-447)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              ||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        141 SerGluAsnProAspAspMetTrp 148
              |||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGG 444

RESULT 4
AX057353
LOCUS       AX057353                326 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 19 from Patent WO0075305.
ACCESSION   AX057353
VERSION     AX057353.1  GI:12310094
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Lalanne,J.L. and Rocher,C.
  TITLE     Novel candida albicans genes and proteins coded by said genes
  JOURNAL   Patent: WO 0075305-A 19 14-DEC-2000;
            HOECHST MARION ROUSSEL (FR)
FEATURES             Location/Qualifiers
     source          1. .326
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      146 a      27 c      53 g     100 t
ORIGIN

Alignment Scores:
Pred. No.:                4.58e-45        Length:         326
Score:                    562.00          Matches:        107
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    99.07%          Mismatches:     0
Query Match:              72.80%          Indels:         0
DB:                       6               Gaps:           0

US-09-980-054A-12 (1-148) x AX057353 (1-326)

Qy          5 AspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGly 24
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db          2 GATATAGATAATGTATTAAATTTAGAAGAAGATCAATATGAATTAGGATTTAAAGAAGGT 61

Qy         25 GlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGA 121

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHis 64
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCAT 181

Qy         65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATT 241

Qy         85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATT 301

Qy        105 LysLysAlaArgAsnLysLeuArg 112
              ||||||||||||||||||||||||
Db        302 AAAAAGGCAAGAAATAAATTAAGA 325

RESULT 5
SCYNL260C/c
LOCUS       SCYNL260C               1104 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome XIV reading frame ORF YNL260c.
ACCESSION   Z71536 Y13139
VERSION     Z71536.1  GI:1302312
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1104)
  AUTHORS   Sen-Gupta,M., Gueldener,U., Beinhauer,J., Fiedler,T. and
            Hegemann,J.H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1104)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XIV sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .1104
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="XIV"
```

```
     CDS             complement(320. .916)
                     /note="ORF YNL260c"
                     /codon_start=1
                     /protein_id="CAA96167.1"
                     /db_xref="GI:1302313"
                     /db_xref="SWISS-PROT:P53846"
                     /translation="MVRNRFIRKMKKNLFKSNHLSYLKSKWKVKITGQIKMDFDNLLN
                     LEEQYYQEGFLEGQNENIKQSFLEGKQYGLQVGFQRFTLLGQMEGLCDVIESYGLHSP
                     TLEKNIHTIRTLMKGLKMNNDDESVMEFERVLIKLKNKFRTILITLHRLVKDKRTPTV
                     TFEVFEDVSRAIAGEIRGFVENEDIAKNKTKQNQAQSW"
BASE COUNT      342 a    206 c    157 g    399 t
ORIGIN

Alignment Scores:
Pred. No.:                2.72e-12      Length:          1104
Score:                    223.00        Matches:         47
Percent Similarity:       58.78%        Conservative:    30
Best Local Similarity:    35.88%        Mismatches:      40
Query Match:              28.89%        Indels:          14
DB:                       8             Gaps:            3

US-09-980-054A-12 (1-148) x SCYNL260C (1-1104)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              :::|||   ||||||:::|||||||||||||||:::   |||:::   ||||||   |||
Db        808 ATGGATTTTGATAATTTACTAAACCTAGAAGAACAGTATTACCAAGAAGGTTTTCTGGAG 749

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              ||||||   :::      |||      :::||||||||||||:::||||||   |||
Db        748 GGCCAGAATGAGAATATAAAGCAGAGCTTTCTAGAAGGTAAACAGTATGGTTTACAAGTG 689

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              ||||||||||||||   :::::|||   ::::::   |||
Db        688 GGATTCCAAAGATTTACGCTCCTGGGACAAATGGAAGGTTTGTGTGATGTT--------- 638

Qy         64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82
              |||:::   |||      :::|||   :::|||      :::::::::   :::
Db        637 ---ATAGAATCGTATGGATTGCATAGTCCAACACTAGAAAAGAATATACATACCATTCGT 581

Qy        ,83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
              :::|||      :::   :::   |||   |||:::  |||   ::::::|||:::
Db        580 ACTTTAATGAAAGGTTTGAAAATGAATAACGATGATGAAAGTGTGATGGAATTTGAGAGG 521

Qy        103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
              :::   |||   :::||||||   |||   |||
Db        520 GTGTTAATAAAATTGAAAAATAAGTTTAGAACCATT---------------------- 485

Qy        123 LysIleAspSerLeuAspAsnLeuValLysGlu 133
              :::   :::|||      |||||||||:::
Db        484 ---CTAATAACTCTTCATCGATTAGTTAAAGAT 455

RESULT 6
YSCATX1HF
LOCUS       YSCATX1HF               2961 bp    DNA     linear   PLN 06-JUN-1995
DEFINITION  Saccharomyces cerevisiae metal homeostasis factor (ATX1) gene,
            complete cds.
ACCESSION   L35270
VERSION     L35270.1  GI:530995
KEYWORDS    metal homeostasis factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2961)
  AUTHORS   Lin,S.J. and Culotta,V.C.
  TITLE     The ATX1 gene of Saccharomyces cerevisiae encodes a small metal
            homeostasis factor that protects cells against reactive oxygen
            toxicity
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (9), 3784-3788 (1995)
  MEDLINE   95249558
   PUBMED   7731983
COMMENT     Original source text: Saccharomyces cerevisiae DNA.
FEATURES             Location/Qualifiers
     source          1. .2961
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     gene            446. .667
                     /gene="ATX1"
     CDS             446. .667
                     /gene="ATX1"
                     /function="protects against oxygen toxicity"
                     /codon_start=1
                     /product="metal homeostasis factor"
                     /protein_id="AAC37428.1"
                     /db_xref="GI:530996"
                     /translation="MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQL
                     VDVYTTLPYDFILEKIKKTGKEVRSGKQL"
BASE COUNT      988 a    466 c    554 g    953 t
ORIGIN

Alignment Scores:
Pred. No.:                8.22e-12      Length:          2961
Score:                    223.00        Matches:         47
Percent Similarity:       58.78%        Conservative:    30
Best Local Similarity:    35.88%        Mismatches:      40
Query Match:              28.89%        Indels:          14
DB:                       8             Gaps:            3

US-09-980-054A-12 (1-148) x YSCATX1HF (1-2961)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              :::|||   ||||||:::|||||||||||||||:::   |||:::   ||||||   |||
Db        969 ATGGATTTTGATAATTTACTAAACCTAGAAGAACAGTATTACCAAGAAGGTTTTCTGGAG 1028

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              ||||||   :::      |||      :::||||||||||||:::||||||   |||
Db       1029 GGCCAGAATGAGAATATAAAGCAGAGCTTTCTAGAAGGTAAACAGTATGGTTTACAAGTG 1088

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              ||||||||||||||   :::::|||   ::::::   |||
Db       1089 GGATTCCAAAGATTTACGCTCCTGGGACAAATGGAAGGTTTGTGTGATGTT--------- 1139

Qy         64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82
              |||:::   |||      :::|||   :::|||      :::::::::   :::
Db       1140 ---ATAGAATCGTATGGATTGCATAGTCCAACACTAGAAAAGAATATACATACCATTCGT 1196

Qy         83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
              :::|||      :::   :::   |||   |||:::  |||   ::::::|||:::
Db       1197 ACTTTAATGAAAGGTTTGAAAATGAATAACGATGATGAAAGTGTGATGGAATTTGAGAGG 1256

Qy        103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
              :::   |||   :::||||||   |||   |||
Db       1257 GTGTTAATAAAATTGAAAAATAAGTTTAGAACCATT---------------------- 1292

Qy        123 LysIleAspSerLeuAspAsnLeuValLysGlu 133
              :::   :::|||      |||||||||:::
Db       1293 ---CTAATAACTCTTCATCGATTAGTTAAAGAT 1322

RESULT 7
SCCHXIVL/c
LOCUS       SCCHXIVL                33016 bp    DNA     linear   PLN 30-JUL-1997
DEFINITION  S.cerevisiae DNA region from chromosome XIV, left arm.
ACCESSION   X96722
VERSION     X96722.1  GI:1255956
KEYWORDS    ATX1 gene; MPA43 gene; NRD1 gene; ORC5 gene; RAD50 gene; RPA49
            gene; SIP3 gene; sui1 gene.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Sen-Gupta,M., Guldener,U., Beinhauer,J., Fiedler,T. and
            Hegemann,J.H.
  TITLE     Sequence analysis of the 33 kb long region between ORC5 and SUI1
            from the left arm of chromosome XIV from Saccharomyces cerevisiae
```

```
  JOURNAL   Yeast 13 (9), 849-860 (1997)
  MEDLINE   97377992
   PUBMED   9234673
REFERENCE   2  (bases 1 to 33016)
  AUTHORS   Hegemann,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-1996) J.H. Hegemann, Institute of Micro &
            Molecular Biology, Justus-Liebig-Univ. Giessen, Frankfurter Str.
            107, D- 35392 Giessen, FRG
FEATURES             Location/Qualifiers
     source          1. .33016
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /sub_strain="FY1679"
                     /db_xref="taxon:4932"
                     /chromosome="XIV"
                     /map="132kb, left arm"
                     /clone="cosmid 14-4c pou"
     gene            223. .1662
                     /gene="ORC5"
     CDS             223. .1662
                     /gene="ORC5"
                     /note="ORF N0834"
                     /codon_start=1
                     /protein_id="CAA65483.1"
                     /db_xref="GI:1255957"
                     /db_xref="SWISS-PROT:P50874"
                     /translation="MNVTTPEVAFREYQTNCLASYISADPDITPSNLILQGYSGTGKT
                     YTLKKYFNAANPNLHAVWLEPVELVSWKPLLQAIARTVQYKLKTLYPNIPTTDYDPLQV
                     EEPFLLVKTLHNIFVQYESLQEKTCLFLILDGFDSLQDLDAALFNKYIKLNELLPKDS
                     KINIKFIYTMLETSFLQRYSTHCIPTVMFPRYNVDEVSTILVMSRCGELMEDSCLRKR
                     IIEEQITDCTDDQFQNVAANFIHLIVQAFHSYTGNDIFALNDLIDFKWPKYVSRITKE
                     NIFEPLALYKSAIKLFLSTDDNLSENGQGESAITTNRDDLENSQTYDLSIISKYLLIA
                     SYICSYLEPRYDASIFSRKTRIIQGRAAYGRRKKKEVNPRYLQPSLFAIERLLAIFQA
                     IFPIQGKAESGSLSALREESLMKANIEVFQNLSELHTLKLIATTMNKNIDYLSPKVRW
                     KVNVPWEIIKEISESVHFNISDYFSDIHE"
     CDS             complement(1982. .2578)
                     /note="ORF N0838"
                     /codon_start=1
                     /protein_id="CAA65484.1"
                     /db_xref="GI:1255958"
                     /db_xref="SWISS-PROT:P53846"
                     /translation="MVRNRFIRKMKKNLFKSNHLSYLKSKWKVKITGQIKMDFDNLLN
                     LEEQYYQEGFLEGQNENIKQSFLEGKQYGLQVGFQRFTLLGQMEGLCDVIESYGLHSP
                     TLEKNIHTIRTLMKGLKMNNDDESVMEFERVLIKLKNKFRTILITLHRLVKDKRTPTV
                     TFEVFEDVSRAIAGEIRGFVENEDIAKNKTKQNQAQSW"
     gene            complement(2767. .2988)
                     /gene="ATX1"
     CDS             complement(2767. .2988)
                     /gene="ATX1"
                     /note="ORF N0840"
                     /codon_start=1
                     /protein_id="CAA65485.1"
                     /db_xref="GI:1255959"
                     /db_xref="SWISS-PROT:P38636"
                     /translation="MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQL
                     VDVYTTLPYDFILEKIKKTGKEVRSGKQL"
     CDS             complement(3232. .5496)
                     /note="ORF N0842"
                     /codon_start=1
                     /protein_id="CAA65486.1"
                     /db_xref="GI:1255960"
                     /db_xref="SWISS-PROT:P53847"
                     /translation="MESLFPNKGEIIRELLKDPLILKNDSKRSNGSELELDSSDLLQR
                     EAILANELNILDNLKTFLNLIKEVKTNLNILELENCYYSLQSLRKKMRNNAAYLKQSF
                     NFQQSISTYVDTLHLELVSTLYKILTNGFWKITENSIQFTPTVEWGKDKVHIEYDTFM
                     DFVAQQYFPKGSLDNQAWFILDMTSADSQEQVRAKLNTIMKEYMNLSRIVSMIKNSIF
                     ISGKEISYENEKNILVFSKSSSHGQHCVSTVLTSFEAVCDFMLDGLAFRDRKTLSYEL
                     GPLFNTEFTKFVKNNASIILESLDSPLKNLVSVINNKLTRLVAKSEVTNWTHSGKEIQ
                     DLLMNKQLYYNLLLDKVLESHISEIRSIFEDPKKSWQNLEVVELTTSNTNTMSEKIGK
                     NDSDVQNEKELHNAVSKDDDWNWEVEDDDADAWGDEIDVNIDDEEEKTNQEKEKEPEE
                     EENAWDEAWAIDENIDDASLENGKEHLKAHDVGSLDKDHIEVTQLPKLFLAISQNFKS
                     SFADSHVDEQYFAYKYNLLQTSYMAMCTANFSHNWCQLYVDMRYLIERDEKLYRIKEL
                     TRNLLETKLNMKYRIVCQLIRHQLTEFRENERNPSWDATIEKLLPYILKEIVRPLQKI
                     RGEEGSRYLLSFLNFLYNDCVTKEILKWQIISEVNSENLGELVSLLVNNTDIQLLAKE
                     PSYKKMREKFATMGKFLPLHLKEIMEMFYNGDFYLFATDELIQWIELLFADTPLRRNA
                     IDDIYEIRGTALDD"
     gene            complement(5754. .9443)
                     /gene="SIP3"
     CDS             complement(5754. .9443)
                     /gene="SIP3"
                     /note="ORF N0844"
                     /codon_start=1
                     /protein_id="CAA65487.1"
                     /db_xref="GI:1255961"
                     /db_xref="SWISS-PROT:P38717"
                     /translation="MSVHGRDPKKRQLRLISVAFKEASIDSPSFRASVNFFQTRVDAL
                     EDWIEKTVDFFDQKYKVSFEDFRRAKETLLSQLLPPPALLSNGFVSNQSFTPRLIDSF
                     NKDYYDFSMKLLQIVKGDDSSHSTALLELMTTAIEPYRNVRKNFDFYQGKYDSMLASY
                     QAIRISKTSLEPSSIKSDALQLFEVQKNYLKASLDLISAISAVKLSLDKFILESMKVL
                     KSRSIFITKDSGRKIDLSPCINEYLDNYAIWVENSIEGSKVLDSDISNAKKQAYRYTL
                     KRITPSSDTSDYNIRSIHSSKLLSKDTQVPPKSPEKSGWLYMKTQVGKPTREIWVRRW
                     CFLKNAVFGMFLLSPSKTYVEETDKFGVFLTNVRYDPEEDRKFCFEVKIFGNKVTEAH
                     DNMSKDITLVFQTSNYLDLKSWLIAFEATKKYVMSIQHDSLEYELAFKRFSPKFFEFA
                     SSTTTSIDQLITTFDKETESLYETLNCSISEYDILTLGEEKVFQFQMPTTPISTKMTQ
                     LAILSNFLTKGSWFPNAVLANIWGTTDWSEYTILPGKGKKPSSLLTIDGKRLPIRNST
                     IYPQYYSNELKVLDLQFKSLVFSPDQRLEKLPEELLLFKFEALWCPNKKQKFSATCFC
                     TKDYIYCYMNSMEFICLTKISLSEIVSVEADRSSKKTLKLYDASGLQMKAIVLFSDYK
                     LIASKLQYLLENKAIKNPNSNEEILVKFEQMEKESQEKKQEELYKIEQENSFDRKATS
                     VSKIIKSRVTFWEMSDDASTLLNRLKKLQTEYSITYNHEYEISSKGLAHILFGDKSNA
                     FPKCLFLARKDGEEHGKRFWYKNKDINGKSQLVRKIPFRLDMTGNFLNTGKYHRDKES
                     KMIFATQRIVKIVDNKYYEVDLDPFFVKVPFCHLLKLSIKFVITESYDVDNHLEIKLN
                     MTASSSSLHVLYKLEYIDSRTGKTIEKLSLAEIICQTWALKFAHSEFLLIRRVLRYYL
                     EKIGKHGKVIKAIKLCGILGVLSNKSEEPATEKNGNSKESESMQYDIRYSCTILFLVF
                     IKLMVYRVTNLTFVFFRILIGILLLCAEKFSRINRMMVVGLLASIMINILLSEKASVP
                     YWSIKRAEKLFHDRLGSDKFTMQRAIYISDSDLLSSQLSVPSNNPIFEKFSEDNFNKD
                     YQYSETRKQLAMRRNELLIELRILQDMEKQLVHDDYEKFLLEEVNKCSMVSIEMTDLW
                     FNDTQLQNYCSICNEELEKLRPPIT"
     CDS             9626. .12220
                     /note="ORF N0848"
                     /codon_start=1
                     /protein_id="CAA65488.1"
                     /db_xref="GI:1255962"
                     /db_xref="SWISS-PROT:P53848"
                     /translation="MINCKVIENRRTRFFPLVNAKLVKVIASSLINKFQSLPLAMSKL
                     FSTVNSARHSVPLGGMRDYVHIKKLEMNTVLGPDSWNQLMPQKCLLSLDMGTDFSKSA
                     ATDDLKYSLNYAVISRDLTNFVSKKKNWGSVSNLAKSVSQFVMDKYSGVECLNLEVQA
                     DTTHIRSDHISCIIQQERGNPESQEFDVVRISELKMLTLIGVFTFERLKKQYVTLDIK
                     LPWPKKAELPPPVQSIIDNVVKFVEESNFKTVEALVESVSAVIAHNEYFQKFPDSPLV
                     VKVLKLNAITATEGVGVSCIREPREIAMVNIPYLSSIHESSDIKFQLSSSQNTPIEGK
                     NTWKRAFLAFGSNIGDRFKHIQMALQLLSREKTVKLRNISSIFESEPMYFKDQTPFMN
                     GCVEVETLLTPSELLKLCKKIEYEELQRVKHFDNGPRTIDLDIVMFLNSAGEDIIVNE
                     PDLNIPHPRMLERTFVLEPLCELISPVHLHPVTAEPIVDHLKQLYDKQHDEDTLWKLV
                     PLPYRSGVEPRFLKFKTATKLDEFTGETNRITVSPTYIMAIFNATPDSFSDGGEHFAD
                     IESQLNDIIKLCKDALYLHESVIIDVGGCSTRPNSIQASEEEEIRRSIPLIKAIREST
                     ELPQDKVILSIDTYRSNVAKEAIKVGVDIINDISGGLFDSNMFAVIAENPEICYILSH
                     TRGDISTMNRLAHYENFALGDSIQQEFVHNTDIQQLDDLKDKTVLIRNVGQEIGERYI
                     KAIDNGVKRWQILIDPGLGFAKTWKQNLQIIRHIPILKNYSFTMNSNNSQVYVNLRNM
                     PVLLGPSRKKFIGHITKDVDAKQRDFATGAVVASCIGFGSDMVRVHDVKNCSKSIKLA
                     DAIYKGLE"
     CDS             complement(12452. .12913)
                     /note="ORF N0852"
                     /codon_start=1
                     /protein_id="CAA65489.1"
                     /db_xref="GI:1255963"
                     /db_xref="SWISS-PROT:P53849"
                     /translation="MSQKACYVCGKIGHLAEDCDSERLCYNCNKPGHVQTDCTMPRTV
                     EFKQCYNCGETGHVRSECTVQRCFNCNQTGHISRECPEPKKTSRFSKVSCYKCGGPNH
                     MAKDCMKEDGISGLKCYTCGQAGHMSRDCQNDRLCYNCNETGHISKDCPKA"
     CDS             complement(13166. .14371)
                     /note="ORF N0856"
                     /codon_start=1
                     /protein_id="CAA65490.1"
                     /db_xref="GI:1255964"
                     /db_xref="SWISS-PROT:P53850"
```

```
                     /translation="MVGPGLGINRVRRKGVYSTKKGSGDNLLLMKRQGKHDIHDRESD
                     DLSGHDAFSPSKKRGKIDSITEDEIEVKKLSTVATFDKLSRSFPNSEVQAAKNAALRG
                     KEKEEEKVVSIPLIQNLKNEDIESIKCRNNNLLDGKKLLLEAELSAVEDNQIFSSSFP
                     EDKKLSLQSCLSSKEQIIKKLQVREEYMSKFKLPPMLFSDELLTEVEPFMPIVMDILE
                     GKISSVYYFEAKNAFKNSQKAYLSVDEFRKLNLNKFTAGFYGLKRQLRVGEEIAKRYK
                     RALTHNQPATLKWWGITDFCNYVLAPETLTSFCIYQLNLSNKSCSSKTPNKHPKQQLN
                     EKEYYYDPELRMLAYDLLEDTVEYGIIVADSDPIEQWEAAIEEDRLRELKLDVHNYSS
                     RRWRLDTHD"
     CDS             15141. .16409
                     /note="ORF N0860"
                     /codon_start=1
                     /protein_id="CAA65491.1"
                     /db_xref="GI:1255965"
                     /db_xref="SWISS-PROT:P53851"
                     /translation="MSTIGAVDILNQKTITSEVAASVTSKYLQSTFSKGNTSHIEDKR
                     FIHVSSRSHSRFTSTPITPNEILSLKFHVSGSSMAYSRMDGSLTVWFIKDASFDKSVE
                     VYIPDCCGSDKLATDLSWNPTSLNQIAVVSNSSEISLLLINEKSLTASKLRTLSLGSK
                     TKVNTCLYDPLGNWLLAATKSEKIYLFDVKKDHSSVCSLNISDISQEDNDVVYSLAWS
                     NGGSHIFIGFKSGYLAILKAKHGILEVCTKIKAHTGPITEIKMDPWGRNFITGSIDGN
                     CYVWNMKSLCCELIINDLNSAVTTLDVCHLGKILGICTEDEMVYFYDLNSGNLLHSKS
                     LANYKTDPVLKFYPDKSWYIMSGKNDTLSNHFVKNEKNLITYWKDMFDNTMIEKRRKN
                     NGGGNNHNKRTSKNTDRIGKDRPSRFNSKK"
     CDS             complement(16564. .17409)
                     /note="ORF N0864"
                     /codon_start=1
                     /protein_id="CAA65492.1"
                     /db_xref="GI:1255966"
                     /db_xref="SWISS-PROT:P36528"
                     /translation="MKVNLMLKRGLATATATASSAPPKIKVGVLLSRIPIIKSELNEL
                     EKKYYEYQSELEKRLMWTFPAYFYFKKGTVAEHKFLSLQKGPISKKNGIWFPRGIPDI
                     KHGRERSTKQEVKLSDDSTVAFSNNQKEQSKDDVNRPVIPNDRITEADRSNDMKSLER
                     QLSRTLYLLVKDKSGTWKFPNFDLSDESKPLHVHAENELKLLSGDQIYTWSVSATPIG
                     VLQDERNRTAEFIVKSHILAGKFDLVASKNDAFEDFAWLTKGEISEYVPKDYFNKTEF
                     LLADN"
     gene            complement(17711. .19438)
                     /gene="NRD1"
     CDS             complement(17711. .19438)
                     /gene="NRD1"
                     /note="ORF N0868"

Alignment Scores:
Pred. No.:                1.23e-10        Length:         33016
Score:                    223.00          Matches:        47
Percent Similarity:       58.78%          Conservative:   30
Best Local Similarity:    35.88%          Mismatches:     40
Query Match:              28.89%          Indels:         14
DB:                       8               Gaps:           3

US-09-980-054A-12 (1-148) x SCCHXIVL (1-33016)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              :::|||   ||||||:::||||||||||||||||:::   |||:::   ||||||   |||
Db       2470 ATGGATTTTGATAATTTACTAAACCTAGAAGAACAGTATTACCAAGAAGGTTTTCTGGAG 2411

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              ||||||   :::      |||      :::||||||||||||:::||||||   |||
Db       2410 GGCCAGAATGAGAATATAAAGCAGAGCTTTCTAGAAGGTAAACAGTATGGTTTACAAGTG 2351

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              ||||||||||||||   ::::::|||   ::::::   |||
Db       2350 GGATTCCAAAGATTTACGCTCCTGGGACAAATGGAAGGTTTGTGTGATGTT--------- 2300

Qy         64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82
              |||:::   |||      :::|||   :::|||        :::::::::   :::
Db       2299 ---ATAGAATCGTATGGATTGCATAGTCCAACACTAGAAAAGAATATACATACCATTCGT 2243

Qy         83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
              :::|||      :::   :::   |||   |||:::   |||   ::::::|||:::
Db       2242 ACTTTAATGAAAGGTTTGAAAATGAATAACGATGATGAAAGTGTGATGGAATTTGAGAGG 2183

Qy        103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
              :::   |||   :::||||||   |||   |||
Db       2182 GTGTTAATAAAATTGAAAAATAAGTTTAGAACCATT---------------------- 2147

Qy        123 LysIleAspSerLeuAspAsnLeuValLysGlu 133
                  :::   :::|||      |||||||||:::
Db       2146 ---CTAATAACTCTTCATCGATTAGTTAAAGAT 2117

RESULT 8
SPCC191
LOCUS       SPCC191                 25000 bp    DNA     linear   PLN 29-MAR-2000
DEFINITION  S.pombe chromosome III cosmid c191.
ACCESSION   AL049644
VERSION     AL049644.1  GI:4678667
KEYWORDS    acetyl-coenzyme a synthetase; AMP-binding, AMP-binding; cyc1;
            cytochrome c; glutathione s-transferase; Glycosyl hydrolase; inv1;
            invertase; LTR; pseudogene.
SOURCE      Schizosaccharomyces pombe (fission yeast)
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 25000)
  AUTHORS   Lyne,M., Rajandream,M.A., Barrell,B.G. and Volckaert,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1998) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and atholieke Universiteit Leuven, Faculty of Agricultural and
            Applied Biological Sciences, Laboratory of Gene Technology,
            Kardinaal Mercierlaan 92 Blok F, B-3001 Leuven, Belgium
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c191 overlaps c417 at the left end and c1450 at the right
            end.
FEATURES             Location/Qualifiers
     source          1. .25000
                     /organism="Schizosaccharomyces pombe"
                     /mol_type="genomic DNA"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="I"
                     /clone="cosmid c191"
     misc_feature    1. .4556
                     /note="nominal overlap with SPCC417 AL035076 S.pombe
                     chromosome III cosmid c417."
     gene            2268. .2804
                     /gene="SPCC191.01"
                     /note="synonym: SPCC417.13"
     CDS             2268. .2804
                     /gene="SPCC191.01"
                     /note="SPCC191.01, len:178"
                     /codon_start=1
                     /label=SPCC191.01
                     /product="hypothetical protein"
                     /protein_id="CAB41047.1"
```

```
                /db_xref="GI:4678668"
                /db_xref="SPTREMBL:O94733"
                /translation="MVTRRLSTNVHIDLQPIRQGILLPFHDRPAEMRDLAKCNDQFFR
                NVRDTISEPKFTQLMELWLEKSRAEVPDADFLMTTRNIMLSFPSNEEQGHQRSASHGS
                TSSATSTPKRLSISSMDPARIHLWRQFCNVVGYDMPTPGEQKEAPPSPAQEAIPESPV
                EEAIASDSDEEEEEEIKA"
gene            complement(4041. .6029)
                /gene="SPCC191.02c"
CDS             complement(4041. .6029)
                /gene="SPCC191.02c"
                /note="SPCC191.02c, len:662, SIMILARITY:Penicillium
                chrysogenum., ACSA_PENCH, acetyl-coenzyme a synthetase,
                (669 aa), fasta scores: opt: 2756, E():0, (61.5% identity
                in 647 aa)"
                /codon_start=1
                /label=SPCC191.02c
                /product="acetyl-coenzyme a synthetase"
                /protein_id="CAB41048.1"
                /db_xref="GI:4678669"
                /db_xref="SPTREMBL:P78773"
                /translation="MTKNPVDHTLIIEPPVRLHGDPTVPKPNIASLDEYKRMYEESIN
                DPSTFWGNMARDMMTWDKQFSTVVQGSIDKADSAWFADGAISPCYNLVDRHAIARPDA
                VALIYEADEPNQGRYITYRELLASVSQCAGALQSMGVGMGDRVAIYMPMIPETIIAML
                AIVRLGAIHSVIFAGFSAESVADRVNDSECKVIITADESHRGGKRIPLKGVVNKALTE
                CPTIKKVLVFQRSAEPTASMVEGRDVWWHDIIPKFPRYCPPAVVNPEHPLFLLYTSGS
                TGKPKGVVHCTGGYLLGAAATCKYVFDLHPTDRMGCAGDVGWITGHTYIVYGPLMLGA
                ATLVFESTPAYPDYSRYWSVVERHRLTQWYIAPTAIRLLQRAGNEFVKHDRSSLRVLG
                SVGEPIAPESFMWYYEVVGEKRCAVADTYWQTETGSHIVTSLGPVTPMKPGSATLPFF
                GIDAVIIDPLTGKIIEGNDVEGVLAIRSPWPSAARTVWRGHDRYIDTYLKPYPGFYFT
                GDGATRDKDGYIWIRGRVDDVVNISGHRLSTAEIEAALLSHDAVAESAVVGVHDELTG
                QAVNAFILLKPGYEATVELEKELIMAVRSTIGPFASPRKLIFSDLPKTRSGKIMRRIL
                RKILAGEVDQIGDLSTLADPKVVEHIIHAVHYAHQKKP"
misc_feature    complement(4359. .5675)
                /gene="SPCC191.02c"
                /note="Pfam match to entry PF00501 AMP-binding,
                AMP-binding enzyme"
gene            complement(6748. .7101)
                /gene="SPCC191.03c"
CDS             complement(6748. .7101)
                /gene="SPCC191.03c"
                /note="SPCC191.03c, len:117, very questionable ORF"
                /codon_start=1
                /label=SPCC191.03c
                /product="very hypothetical protein"
                /protein_id="CAB41049.1"
                /db_xref="GI:4678670"
                /db_xref="SPTREMBL:Q9Y7P7"
                /translation="MRSNNSSLVHCCWVSPPSLTRLPAFPSPRILSPCYCYNKRIRPF
                RGLTSYRQASYSLGFPLGLLVFLHSLIVARFFVASKSRSCIVRSLLFWINLDSADSRI
                SVLFQCFFCIDIWTV"
gene            complement(7329. .7631)
                /gene="SPCC191.04c"
CDS             complement(7329. .7631)
                /gene="SPCC191.04c"
                /note="SPCC191.04c, len:100, questionable ORF"
                /codon_start=1
                /label=SPCC191.04c
                /product="hypothetical protein"
                /protein_id="CAB41050.1"
                /db_xref="GI:4678671"
                /db_xref="SPTREMBL:Q9Y7P8"
                /translation="MHSVCSIFLSCSHRVIQAKHPPFPLFHSYFHIPDFLSFVFPFVA
                SPPLAFARRKLDHVPKKFARSIGPFLLIVFLFFNLFPTFFFLPFFPDTTKRPNLAD"
gene            complement(8126. .8761)
                /gene="SPCC191.05c"
CDS             complement(8126. .8761)
                /gene="SPCC191.05c"
                /note="SPCC191.05c, len:211"
                /codon_start=1
                /label=SPCC191.05c
                /product="hypothetical protein"
                /protein_id="CAB41051.1"
                /db_xref="GI:4678672"
                /db_xref="SPTREMBL:Q9Y7P9"
                /translation="MTGTAKVSIPKAQPHKVPCVYLAGDLVFRPNAIELFDELKEICK
                DAGVQGVAPFDGQEGVEEMAPGAETSLKIAELDRKLMDRCDGGIFCLDPFRRAPDMDP
                GTAVELGYMAAQGKPLAGFTTDGRMYPEKVRSYRKQAWGDALKPRFTKGGSGSMEDAD
                GLIVHSEGFLQNVMTEGFIRMSGGFVAVDFSIQEAFRTAIKDLAARLNSQH"
gene            9946. .10362
                /gene="SPCC191.06"
CDS             9946. .10362
                /gene="SPCC191.06"
                /note="SPCC191.06, len:138, questionable ORF"
                /codon_start=1
                /label=SPCC191.06
                /product="very hypothetical protein"
                /protein_id="CAB41052.1"
                /db_xref="GI:4678673"
                /db_xref="SPTREMBL:Q9Y7Q0"
                /translation="MSHLFILSCSVYDFSSILLHQYFTQFFVLRYASPSMHLDWRFPS
                LRIHSYPKDQSTHAWKMHCMPTEIIAYQPIPVFQRLLHSNDIYFPYFQTRYTTLIMFN
                TVCSSSPNPVFTLTKVIARKKGKHWIGSPCHSAESS"
gene            10542. .10871
                /gene="SPCC191.07"
                /note="synonym: cyc1"
CDS             10542. .10871
                /gene="SPCC191.07"
                /note="SPCC191.07, len:109"
                /codon_start=1
                /label=cyc1
                /product="cytochrome c."
                /protein_id="CAB41053.1"
                /db_xref="GI:4678674"
                /translation="MPYAPGDEKKGASLFKTRCAQCHTVEKGGANKVGPNLHGVFGRK
                TGQAEGFSYTEANRDKGITWDEETLFAYLENPKKYIPGTKMAFAGFKKPADRNNVITY
                LKKATSE"
misc_feature    10557. .10862
                /gene="SPCC191.07"
                /note="Pfam match to entry PF00034 cytochrome_c,
                Cytochrome c"
LTR             complement(12088. .12436)
                /note="tf2-type LTR"
gene            12558. .12962
                /gene="SPCC191.08"
CDS             12558. .12962
                /gene="SPCC191.08"
                /note="SPCC191.08, len:134"
                /codon_start=1
                /label=SPCC191.08
                /product="hypothetical protein"
                /protein_id="CAB41054.1"
                /db_xref="GI:4678675"
                /db_xref="SPTREMBL:Q9Y7Q1"
                /translation="MDWEEVTKLEENEYKRGYDEGILKGIEQGYEEAFLFGLEHAYNK
                YLLAGEIYGRVCFWLKEENSQHPKIKKAHRHLEQLKSLLESLPTNNELEETDAGFDSY
                WNKITAKAKVVSSLLGTKILPAEKIDANDGFE"
gene            complement(15132. .15821)
                /gene="SPCC191.09c"
CDS             complement(15132. .15821)
                /gene="SPCC191.09c"
                /note="SPCC191.09c, len:229,
                SIMILARITY:Schizosaccharomyces pombe, O59827, glutathione
                s-transferase., (230 aa), fasta scores: opt: 1220, E():0,
                (78.8% identity in 226 aa)"

Alignment Scores:
Pred. No.:              0.0391          Length:         25000
Score:                  132.00          Matches:        37
Percent Similarity:     50.00%          Conservative:   29
Best Local Similarity:  28.03%          Mismatches:     58
Query Match:            17.10%          Indels:         8
DB:                     8               Gaps:           5

US-09-980-054A-12 (1-148) x SPCC191 (1-25000)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
```

```
Db      12558 ATGGACTGGGAGGAAGTGACCAAATTGGAAGAAAATGAGTACAAACGTGGTTATGATGAA 12617

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43

Db      12618 GGAATTTTAAAGGGAATTGAGCAAGGGTATGAGGAAGCATTTTTGTTTGGACTCGAGCAT 12677

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63

Db      12678 GCTTATAACAAATATTTACTGGCGGGTGAAATTTATGGACGCGTGTGTTTTTGGTTGAAA 12737

Qy         64 HisIleAsp---GlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82

Db      12738 GAAGAAAATTCACAACATCCTAAAATTAAAAAGGCACATAGGCATTTAGAACAGTTGAAA 12797

Qy         83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAsp------Tyr 100

Db      12798 TCGTTATTAGAGTCGTTGCCC---ACCAACAATGAACTGGAGGAGACCGATGCTGGGTTT 12854

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle------Thr 118

Db      12855 GACTCTTACTGGAACAAAATTACAGCCAAAGCCAAAGTGGTATCGAGTCTTTTAGGAACT 12914

Qy        119 Lys------GluThrTrpLysIleAspSerLeuAsp 128

Db      12915 AAGATTCTTCCCGCTGAAAAAATTGATGCTAATGAT 12950
```

```
RESULT 9
AP003953
LOCUS       AP003953               105692 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OJ1365_D04, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP003953
VERSION     AP003953.1  GI:15021923
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1365_D04
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 105692)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nuclectide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .105692
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="OJ1365_D04"
```

```
BASE COUNT    30807 a  22247 c  23067 g  29571 t
ORIGIN

Alignment Scores:
Pred. No.:                0.305          Length:          105692
Score:                    130.00         Matches:         36
Percent Similarity:       52.14%         Conservative:    25
Best Local Similarity:    30.77%         Mismatches:      49
Query Match:              16.84%         Indels:          7
DB:                       2              Gaps:            3

US-09-980-054A-12 (1-148) x AP003953 (1-105692)

Qy         12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31

Db      37741 TTAGATGAGACACACTATCAAACGGGTTTCAAGAATGGTTATAGTGAGGGCTTGGTGTCT 37800

Qy         32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIle 51

Db      37801 GGAAAAGAAGAGGGAAGGCAGGTTGGTTTAAAGAATGGTTTTCAGGTAGGTGAAGAACTA 37860

Qy         52 GlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------HisIleAspGlnTyrAsn 69

Db      37861 GGTTTTTATCAGGGATGTCTGGATGTTTGGACGTCGTTGGTTTCAATTGATCAAGATGCA 37920

Qy         70 AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSer 89

Db      37921 TTCTCAGCTCGGGTCAGGAAAAACATTGAGCAACTAGCTGCACTGTTGAGAAGCTATCCG 37980

Qy         90 IleThrAsn---GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108

Db      37981 TTGTCGAACCCAGAAGACGAGCAAGTTCAAGATATA---------ATGGAGAAGATAAGA 38031

Qy        109 AsnLysLeuArgValIleAlaSer-IleThrLysGluThrTrpLysIle 124

Db      38032 CTGAAATTCAGGGTTATCACAGCAAGTTTAGGTACAAAACTGGAGTATC 38080
```

```
RESULT 10
AP004674
LOCUS       AP004674               144741 bp    DNA     linear   PLN 19-DEC-2002
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
            PAC clone:P0681F05.
ACCESSION   AP004674
VERSION     AP004674.2  GI:26017274
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
            clone:P0681F05
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 144741)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Dec 3, 2002 this sequence version replaced gi:18307752.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and
            BLASTX. The genomic sequence was searched against NCBI NonRedundant
            Protein database, nr
```

(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
This sequence of P0681F05 clone has an overlap with OSJNBa0007H12 (DDBJ: AP004990) clone at 5' end and with OSJNBb0044C15 (DDBJ: AP005850) at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES Location/Qualifiers

source 1. .144741
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="P0681F05"

gene 3263. .3469
/gene="P0681F05.1"

CDS 3263. .3469
/gene="P0681F05.1"
/note="hypothetical protein
similar to Oryza sativa chromosome7, OJ1656_E11.24"
/codon_start=1
/protein_id="BAC45137.1"
/db_xref="GI:27261021"
/translation="MADGYRRRVLTAASTLVPAACSSAGGMQCVPAARGGGRKLFGVA
TGGDRQLRCGWPRKAVESAFQMAD"

gene complement(join(6526. .6561,6842. .6925,7110. .7217,
7735. .7831,7975. .7985))
/gene="P0681F05.2"

CDS complement(join(6526. .6561,6842. .6925,7110. .7217,
7735. .7831,7975. .7985))
/gene="P0681F05.2"
/note="hypothetical protein
predicted by FGENESH etc."
/codon_start=1
/protein_id="BAC45138.1"
/db_xref="GI:27261022"
/translation="MTIWKPMSLHQSQDMPDGSAFGGARNDWVCMCEFTSQRRYIAEK
SRAAPGPSSHGLASAPCEMPSPPCKHMQGPVLIFMFPKPDRGTRYIGSNPLSGWEQVQ
MRTPTRFMS"

gene complement(join(13012. .13105,13319. .13446))
/gene="P0681F05.3"

CDS complement(join(13012. .13105,13319. .13446))
/gene="P0681F05.3"
/note="hypothetical protein
predicted by GeneMark.hmm etc."
/codon_start=1
/protein_id="BAC45139.1"
/db_xref="GI:27261023"
/translation="MTDAGEGTDGEEPESEERAAARDRSDDESSAAPRLASPPPELGL
GGLSRQPNIRNLHGLVRANLGEGNQTDPT"

gene 14320. .16593
/gene="P0681F05.4"

misc_feature 14320. .16593
/gene="P0681F05.4"
/note="probably inactive due to stop codon(s) in CDS
pseudogene, transposase"
/pseudo

gene join(19824. .19909,20071. .20178,20294. .20405)
/gene="P0681F05.5"

CDS join(19824. .19909,20071. .20178,20294. .20405)
/gene="P0681F05.5"
/note="hypothetical protein
predicted by GeneMark.hmm etc."
/codon_start=1
/protein_id="BAC45140.1"
/db_xref="GI:27261024"
/translation="MGTEELVGASGGDAEAGADEQRRPVAEARRAATGPGGDEARTSN
GGQSRTRRRPVAEANADERRSGAEREAVGADDERRRSVAEAVADEMTAGRGGGGRRG"

gene join(23817. .24701,24780. .24953)
/gene="P0681F05.6"

CDS join(23817. .24701,24780. .24953)
/gene="P0681F05.6"
/note="contains ESTs
D41499(S4032),AU094265(E3758),AU033141(S4032)"
/codon_start=1
/product="putative thiamine biosynthetic enzyme"
/protein_id="BAC45141.1"
/db_xref="GI:27261025"
/translation="MATTASSLLKTSFAGARLPAAARNPTVSVAPRTGGAICNSISSS
SSTPPYDLNAIRFSPIKESIVSREMTRRYMTDMITYADTDVVVVGAGSAGLSCAYELS
KDPSVSVAVIEQSVSPGGGAWLGGQLFSAMVVRKPAHLFLDELGVAYDEQEDYVVIKH
AALFTSTVMSRLLARPNVKLFNAVAVEDLIVKEGRVGGVVTNWALVSMNHDTQSCMDP
NVMESRVVVSSCGHDGPFGATGVKRLQDIGMIDAVPGMRALDMNTAEDEIVRLTREVV
PGMIVTGMEVAEIDGAPRMGPTFGAMMISGQKAAHLALKALGRPNAIDGTIKKAAAAA
AHPELILASKDDGEIVDA"

gene complement(join(27163. .27332,27442. .27529))
/gene="P0681F05.7"

misc_feature complement(join(27163. .27332,27442. .27529))
/gene="P0681F05.7"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GENSCAN"

gene join(28697. .28907,28986. .29086,29178. .29380,29774. .30188)
/gene="P0681F05.8"

CDS join(28697. .28907,28986. .29086,29178. .29380,29774. .30188)
/gene="P0681F05.8"
/note="contains ESTs C72892(E2426),AU172540(E2426)"
/codon_start=1
/product="putative xyloglucan endotransglycosylase"
/protein_id="BAC45142.1"
/db_xref="GI:27261026"
/translation="MGSLGRRPWVGGLTAAMIFAVAVCGFCFSGASAAAAAPTFGDNF
EITGAEDHVKTSADGQTWYLYLDNKTGVGFQTKERYLFGWFSMNLKLAGNDSAGVVTA
YYMCSDVDAAPQRDELDFEFLGNRTGEPYIIQTNVYRSGVGGREMRHSLWFDPTADFH
SYSILWNPKQIVFFVDKVPIREYRNSDKPNTFFPIMKPMYVFSSIWNADDWATRGGLE
KTDWTKAPFISSYRDFTADACSWGTAAASPPSCAASTGNSWWDQPPAWALDAGQREDS
AWVARNLVIYDYCDDRKRFPSPPEECLLRTTSS"

gene complement(join(32633. .32677,32784. .32837,32929. .33140,
33226. .33262))
/gene="P0681F05.9"

CDS complement(join(32633. .32677,32784. .32837,32929. .33140,
33226. .33262))
/gene="P0681F05.9"
/note="contains ESTs
AU093360(C1467),AU075849(S20385),D28232(C1467),
AU075848(S20385)"
/codon_start=1
/product="translational initiation factor eIF1"
/protein_id="BAC45143.1"
/db_xref="GI:27261027"
/translation="MSDLDIQIPTAFDPFAEANAGDSGAAAGSKDYVHVRIQQRNGRK
SLTTVQGLKKEPSYNKILKDLKKEFCCNGTVVQDPELGQVIQLQGDQRKNVSNFLVQA
GIVKKEHIKIHGF"

gene join(36723. .36827,37444. .37566)
/gene="P0681F05.10"

misc_feature join(36723. .36827,37444. .37566)
/gene="P0681F05.10"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GlimmerM"

gene 38236. .38406

```
                     /gene="P0681F05.11"
     misc_feature    38236. .38406
                     /gene="P0681F05.11"
                     /note="This category is not included in IRGSP standard.
                     hypothetical ORF
                     predicted by GlimmerM"
     gene            complement(join(38941. .39014,39099. .39252,40484. .40510))
                     /gene="P0681F05.12"
     CDS             complement(join(38941. .39014,39099. .39252,40484. .40510))
                     /gene="P0681F05.12"
                     /note="hypothetical protein
                     predicted by GlimmerM etc."
                     /codon_start=1
                     /protein_id="BAC45144.1"
                     /db_xref="GI:27261028"
                     /translation="MKPSVKPPEFLTQNNLAPGKEIKTKLVTLCWKSDEEKEQSTEII
                     RLPVPYQLPPKPYGTEDVPWSWDKRYTKKDVYGSVWPRHG"
     gene            complement(join(44510. .44803,44881. .44957,45876. .45936,
                     46437. .46655,47686. .47958,48064. .48255,48513. .48761))
                     /gene="P0681F05.13"

Alignment Scores:
Pred. No.:              0.434          Length:         144741
Score:                  130.00         Matches:        36
Percent Similarity:     52.14%         Conservative:   25
Best Local Similarity:  30.77%         Mismatches:     49
Query Match:            16.84%         Indels:         7
DB:                     8              Gaps:           3

US-09-980-054A-12 (1-148) x AP004674 (1-144741)

Qy          12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
               |||:::|||      |||:::   |||||||||   |||       :::|||
Db       50294 TTAGATGAGACACACTATCAAACGGGTTTCAAGAATGGTTATAGTGAGGGCTTGGTGTCT 50353

Qy          32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIle 51
                  ||||||::::::   |||   :::   |||||||||           :::
Db       50354 GGAAAAGAAGAGGGAAGGCAGGTTGGTTTAAAGAATGGTTTTCAGGTAGGTGAAGAACTA 50413

Qy          52 GlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------HisIleAspGlnTyrAsn 69
               |||:::   |||       :::      |||   |||         |||||||||
Db       50414 GGTTTTTATCAGGGATGTCTGGATGTTTGGACGTCGTTGGTTTCAATTGATCAAGATGCA 50473

Qy          70 AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSer 89
                  |||:::   :::|||   ::::::      |||       ::::::
Db       50474 TTCTCAGCTCGGGTCAGGAAAAACATTGAGCAACTAGCTGCACTGTTGAGAAGCTATCCG 50533

Qy          90 IleThrAsn---GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108
               ::::::|||      |||::::::|||:::|||           ::::::|||   |||
Db       50534 TTGTCGAACCCAGAAGACGAGCAAGTTCAAGATATA---------ATGGAGAAGATAAGA 50584

Qy         109 AsnLysLeuArgValIleAlaSer-IleThrLysGluThrTrpLysIle 124
                  |||   |||||||||   ::: :::     :::   |||   |||
Db       50585 CTGAAATTCAGGGTTATCACAGCAAGTTTAGGTACAAAACTGGAGTATC 50633

RESULT 11
NC20H10
LOCUS       NC20H10                96283 bp    DNA     linear   PLN 12-MAR-2003
DEFINITION  Neurospora crassa DNA linkage group V Cosmid contig 20H10.
ACCESSION   BX294024
VERSION     BX294024.1  GI:28950125
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 96283)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of cosmid 20H10 from 1 to: 21109 and BAC
            12H6 from 21110 to: 96283; cosmid clone 20H10 (cosmid library
            pLORIST6Xh, strain 74-OR-23-1A) and BAC clone 12H6 (strain OR74A)
            are available at the Fungal Genetic Stock Center,
            http://www.fgsc.net Sequencing was performed by MWG Biotech AG,
            Ebersberg, Germany, http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .96283
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="2"
     gene            889. .2585
                     /gene="20H10.010"
     gene            complement(join(889. .1168,1306. .2585))
                     /gene="20H10.010"
     CDS             complement(join(889. .1168,1306. .2585))
                     /gene="20H10.010"
                     /note="match of 36 ESTs
                     similarity to ACYL-COA DEHYDROGENASE, LONG-CHAIN SPECIFIC
                     PRECURSOR, (EC 1.3.99.13), Sus scrofa, SWISSPROT:ACDL_PIG
                     Contains Acyl-CoA dehydrogenases signatures
                     [AITEAFAGSDVAG];Cytochrome b5 family, heme-binding domain
                     signature [FLDAHPGG]
                     contains EST gb:AI328730, BG278778"
                     /codon_start=1
                     /product="related to ACYL-COA DEHYDROGENASE, LONG-CHAIN
                     SPECIFIC PRECURSOR"
                     /protein_id="CAD70984.1"
                     /db_xref="GI:28950126"
                     /translation="MTNVLNVPDKTFTREEVKKHNTEDSTWFIIDSVVYDVSEFLDAH
                     PGGEAVLRQVAGQDATSAFYNLHRHEVLIKYNDLAIGSIKGEKQSIITPRPGDLSPVP
                     YAEPLWLTPQFKSPYFNDSHRSLQKAIRKFTDEVIAPEALEKERTGELISQDLIDKMS
                     KAGILHMRMGPGKHLHGVNLLDGAVKGEEFDYFHDMIMAQEQVRVACRGFQDGNMAGM
                     AIGLTCILNYGDKNPALKKRITEEVFAGKKKICLAITEAFAGSDVAGLRTTATKTPDG
                     KHYIVNGTKKWITNGVFCDYFVTGVKTDKGLSVLLIERGEGVETKPIKTSYSPAAGTT
                     YITFDNVKVPVENLLGEENKGIYVILSNFNHERWMMACATIRYMRLVVEESLKWSHQR
                     LVFGKRLIDQPVIRQKLAKMIALCESHQSWLETITYQMNHMPYSEQAKHLGGPIALLK
                     MSCTRAAHEIADESVQIWGGRGLTQTGMGRVIENFNRGYKFDAILGGAEEVLGDLGVR
                     QAMKFMPKAVL"
     exon            complement(889. .1168)
                     /gene="20H10.010"
                     /number=1
     intron          complement(1169. .1305)
                     /gene="20H10.010"
                     /number=1
     exon            complement(1306. .2585)
                     /gene="20H10.010"
                     /number=2
     gene            4841. .6538
                     /gene="20H10.020"
     CDS             join(4841. .5417,5469. .6538)
                     /gene="20H10.020"
                     /note="similarity to ANON-37CS PROTEIN, Drosophila
                     melanogaster, SWISSPROT:A37C_DROME"
                     /codon_start=1
                     /product="related to ANON-37CS PROTEIN"
                     /protein_id="CAD70985.1"
                     /db_xref="GI:28950127"
                     /translation="MDSTKRPHVGIVGAGMAGLRSAGYLLELGFQVTILEARDRLGGR
                     IYQEKLPNGHFVDMGANWIHGTKENSIFQLAKETGTIATNWDGDAAVFDEHGDILPAK
                     DSERYSTIMWNIIAEAFQYSDKHSADIDSNRSLLDFFKEKVVEQIPETEEDYARKRKI
                     VLQMAELWGAFVGSPVEKQSLKFFWLEECLDGENLFCSGTYRKIMEKIVAPVVDGGAD
                     IKLQTRVAEIFGKSSTGSSNTVKVKTSDNQYYEFEELVLTTPLGWLKQNLQVFHPPLP
```

```
                PRLTTAIQSIGYGCLEKVYISFPKAFWLEPDAYNNNNNNNSNSNKNNRTVKGFCQWLS
                PSYASATNPSRWTNEIVELGSIDPSVAHPTLLFYIYGAESEYVTSKVSSLSSSSAPSS
                EFDADKNGSSPRGQNEATAAAQQFLYDFFKPYYSLLPSYSPTDPNCQPSGCLATDWLH
                DDLAGNGSYCNFQVGLQEGDKDILAMRHGVPEEGVWMAGEHTATFVALGTVTGAYWSG
                EDVARRIAEGYGRGGVAGAGAGAGAGAGAGAGEREAAV"
exon            4841. .5417
                /gene="20H10.020"
                /number=1
intron          5418. .5468
                /gene="20H10.020"
                /number=1
exon            5469. .6538
                /gene="20H10.020"
                /number=2
repeat_region   5817. .5840
                /note="24 bp acaacaacaaca tandem repeat"
repeat_region   6463. .6516
                /note="54 bp tgctggtgctggtgctggtgctggtgctggtgctggtgctgg
                tandem repeat"
gene            9025. .15161
                /gene="20H10.030"
CDS             join(9025. .9078,9111. .15074,15144. .15161)
                /gene="20H10.030"
                /note="similarity to sporulation protein SPO72, YNL242w,
                Saccharomyces cerevisiae, PIR:S63208"
                /codon_start=1
                /product="related to sporulation protein SPO72"
                /protein_id="CAD70986.1"
                /db_xref="GI:28950128"
                /translation="MLDLFYPYVSTPTFHSAFKLEALTNLPPTFKIQKAKVLLLRITI
                PVDIYNSPIIVEVEGVDTRIRVASKQEQDEQTAKNKKGTHKDGDEVVPTAADLAQSFL
                QTEIPTDEQRRLEKALAEEAQEALSESMSDSETDDDGSFATFGTGQGLSLPTFLTNFL
                QGILDRLQIRIHKVTFQLDMQVPVEPGLSTIELVTFQLALDEVIAEGVTAPGQQKDGS
                PAIVPREGKRHILLDNLRAFLITEANVFSSLLPTQSAQSLVKPQGPAPHDVPSTLRDM
                SGSMQSSVGGLDMSISPDQMEALDMLAQSQYEVRDSEDALNIPYESDTQYPEAEENVA
                GSSPLSTPRASMYQEHDTPMLQDHANSAIMQHQPELWSSYERGIRSEPSLHPPTSFQP
                QTMPSGAVSPAPSEPSSSASSVRSDDDTPSADTEDLAQSHLYSHEEAESMYMSAFSDA
                HSTLMPGAYMPGGWGAESEGGESASERDLSTTHPVAGEPQSSTPVVSELASPTPIPEQ
                NPTAQGYEPQHEDTSTPRGITRMVKEIVSLDSISVYVPSTRKQVLAPVTNLAKSNPNL
                PGAFSVHQSFHSSTTDQSTLLTPENLKEDRQDYAIEIVLKPLTLRFDSSVGFLLAMVV
                TRLLEAFKGSSDEGAEVKPSSTESILPDIKVCLEQVTMQFLEELTGVADSAKRIYETQ
                KPNFGSDVLLEASLTNLNAFTHQSGSQTEVNVSIEKFAFGYADEAIVSFTGEADLFQS
                TATVDMLSVGKDIAVKATITPDVTRVDVKTLPLYLKLDLQRLDETFSWFGGLSGFLNM
                GASEDSVAKAVQIPAKPPQKTRGVRFETPIDPMEKRSGSKIDVRINGVHVDVIGKDCR
                AVLNTSALKLVHREEAIAAVLKTIRLSGPYVRNSHARAPVVLELDSTRLDYILTPRTR
                DLERLLELITPSKVKFDEDEDEIMVDTLLRQRRKGGVLGLEVKNFKVRAGNLALLDCI
                PSLVDDLAKLGTVAKYLPEDDRPGLLTLGQVTNVDCEVDVGGRFGVVNARLTNLELAQ
                ITVPSLVAVAVGVVTVNRNENSPIEEELVSTSTAHAPGSSKAPVLMMRMIDDIEPVLK
                IKLFGLNVDYRVPTIMDVLGLIQEETTPEEFEANLAASVANLGGQATAALRRQDTPES
                PVVDKEFKPIKLDVAFRDCVVGLNPLGQDSKLAIVLADSHLEAIPGKDSTLDATATLK
                KACILLIDDVKTLQDVQINARGRAPAMSTPQALELCSKGFVSICEMSSAKAVVKVGKD
                ENGESHVEVSVRDDLLVLETCADSTQTLISLANALTPPTPPSKEIKYRTSVLPVEDLF
                ASITPDAFGRAEGEYDFDDDFAGAQGIECGNEDDDDYYGIGSTEHLEIQSEGYGVAEE
                LFDATNSSLLGDIEVENTNDGMLVSTANLDVPPVSSQSESDLDIQENYFSSEPVKNTT
                LRWNSRKNLYDQSSDAQVFKSPLVICVRDVHVIWNLYDGYDWVRTREIITKAVQDVEA
                KAYERKARADRHTYEDEGDEESVVDDCLFNSIYIAVGPNGDPSNLRRAINQELQYQDT
                TTETESVATTAFTTSTVRASGHRQSRPRGKTLKLSRSKNHKITFELKGVNIDVVTFPP
                GNETINTIDVRIHDLDVFDHIRTSTWKKFAMYDIDAGERELSKHMVHLEVLNVKPVPD
                LPATELVVKVNILPLRLHVDQDALDFITRFFEFKDETQPIHQSSSDVPFIQRCEVGDV
                PVRLDFKPKRVDYAGLRSGHTTEFMNFVILEDSRLVLRHVILYGVSGFDKLGKKLNDI
                WTADVKSTQLPGVLAGVAPVRSLVNAGSGFKDLIEIPIREYKKDGRIFRSIGKGATAF
                AKTTGTEVVKLGAKLAIGTQYALQGAEGMLAKNPPNYNHAGPSSSSAGVPAGVDYEVW
                DEEDFGDHHTPKAISLYADQPLGIMQGMRGAYASLSRDIAIARDAIIAVPTEVMESSS
                AQGAAKAVLMQAPTILFRPAIGVSKAIGQTLLGATNALDPNHRKRIEAKYKKH"
exon            9025. .9078
                /gene="20H10.030"
                /number=1
intron          9079. .9110
                /gene="20H10.030"
                /number=1
exon            9111. .15074
                /gene="20H10.030"
                /number=2
intron          15075. .15143
                /gene="20H10.030"
                /number=2
exon            15144. .15161
                /gene="20H10.030"
                /number=3
repeat_region   15540. .15563
                /note="24 bp catcatcatcat tandem repeat"
repeat_region   15667. .15694
                /note="28 bp cagacagacagacaga tandem repeat"
tRNA            15856. .15927
                /product="tRNA-Arg"
                /note="tRNA predict as a tRNA- Arg : anticodon acg"
gene            16532. .17916
                /gene="20H10.040"
gene            complement(join(16532. .17260,17353. .17631,17704. .17916))
                /gene="20H10.040"
CDS             complement(join(16532. .17260,17353. .17631,17704. .17916))
                /gene="20H10.040"
                /note="similarity to putative fatty acid desaturase mld,
                Homo sapiens, TREMBL:HSAF2668_1
                Contains Neutral zinc metallopeptidases, zinc-binding
                region signature [LAIHEISHNL]"
                /codon_start=1
                /product="related to putative fatty acid desaturase (mld)"
                /protein_id="CAD70987.1"
                /db_xref="GI:28950129"
                /translation="MAASAVTTATTDSSSTQRKQQQNGKLDATSSSSASAGAGDQNDA
                LEKDFFWTYTEEPHRTRRMAIIKAHPEVTKLCGPEPLTKWVVLGVVSLQFLLAHLLRD
                TSFFSWKFWLTAYVFGATCNQNLFLAIHEISHNLAFRSPTANRLFAIVANLPIAVPYS
                ASFRPYHLTHHKSLGVDGLDTDLPTAFEALFLDSILGKAFFCTFQIFFYALRPMAIYR
                VPFTWVHWLNLTIQLFFDYAIVFLLPDSLFTANSLLYFLLSSFLAGSLHPTAGHFIAE
                HYVYEKITPEARNPDNKIPVPETFSYYGPLNWVTYNVGLHNEHHDFPAIPWTRLPKLY
                EIASEFYEGLPQHRSWTHVLWQFIFDEEIGMRCRVKRKVGGRVVGGGTVSAKQQQPQN
                QAKEAGWGDDEIEA"
exon            complement(16532. .17260)
                /gene="20H10.040"
                /number=1

Alignment Scores:
Pred. No.:                1.42          Length:          96283
Score:                    122.50        Matches:         38
Percent Similarity:       32.59%        Conservative:    35
Best Local Similarity:    16.96%        Mismatches:      54
Query Match:              15.87%        Indels:          97
DB:                       8             Gaps:            4

US-09-980-054A-12 (1-148) x NC20H10 (1-96283)

Qy          5 AspIleAspAsn---ValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              |||   :::|||   :::|||||||||||||||||:::   |||:::   |||:::   :::
Db      24044 GACTTTGAAAACTCCCTCCTCAACCTCGAAGAAGATTTCTACCAACAAGGCTACGCCCAA 24103

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              |||         |||   |||           ||||||::::::   |||   :::
Db      24104 GGCGTCTCGGACGGCGCCAAGGCCGGCCGCACCGAAGGCCGCCAGCTCGGCTTGGAAAAG 24163

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              |||||||||:::|||:::         |||  :::                |||   :::
Db      24164 GGGTTCCAGAAGTTTGTCCAGAGCGGCCGGTTGCATGGCCGAGCCATTGTCTGGGCGAAT 24223

Qy         64 HisIle---------------------------------------------------- 65
                 :::
Db      24224 CGGTTACCGTCCAAGCAGGACGTTCGTCGTCATCATCTCCTTCAGCAGCAGCAGCAGCAG 24283

Qy         65 ---------------------------------------------------------- 65

Db      24284 CAGCAGCAGCAGCAGCAGCAGCAGCAGTCATCACAAACCGGCCAACAACAGCAACAACCT 24343

Qy         66 ----------------------AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeu 78
                                    |||       :::||||||       ::::::
Db      24344 CAGAGTTTGAGTCAGAATCAGAGCCAGAGTCAGAGTTCGAGTCAGAGTCAGGACAAGAGT 24403
```

```
Qy          79 AsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGly---------- 93
               :::|||||| :::  :::::::::  |||
Db       24404 CAAAAAGTGGAGGATAGCCTAGAAAACTTAACTCTGGGTAACAACAAGCCGGAGAAAACA 24463

Qy          93 ------------------------------------------------------------ 93

Db       24464 GAGGAGCTGCGCCTACTGCCCAACAACGCCCGTCTAGAAAAACATGTCTCTACCCTCTTT 24523

Qy          94 ----------------------------------------AspLysGluValGluAspTyr 100
                                                       |||:::  |||  |||:::
Db       24524 GCGCTGGTCGAGACCGACACGCTAGCGAAAGAAAACACGGACGAGGCAGTGAATGATTTT 24583

Qy         101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
               :::  :::|||:::|||:::  |||  :::::::::
Db       24584 GATGATCGACTGAAGAGGGCGCAGGGAAAAGCCAAGATTGTG---------------- 24625

Qy         121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
                         :::  :::|||  ||||||      :::
Db       24626 --------------------GAGAGGATGGTGGGGGAGGTGTCAGAAAAGAAGAAGGCG 24664

Qy         141 SerGluAsnPro 144
               :::  ||||||
Db       24665 AATACGAATCCT 24676
```

RESULT 12
AF360255

```
LOCUS       AF360255               1904 bp    mRNA    linear   PLN 18-SEP-2002
DEFINITION  Arabidopsis thaliana unknown protein (At2g20830) mRNA, complete
            cds.
ACCESSION   AF360255
VERSION     AF360255.1  GI:13430685
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1904)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1904)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
            Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
            Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Chen,H.,
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to Genbank.
FEATURES             Location/Qualifiers
     source          1. .1904
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL09-16-J24 (R09231)"
                     /note="This clone is in a modified pBluescript vector
                     (FLC-1) as a BamHI/XhoI insert.
                     ecotype: Columbia"
     gene            1. .1904
                     /gene="At2g20830"
     5'UTR           1. .832
                     /gene="At2g20830"
     CDS             833. .1726
                     /gene="At2g20830"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAK25965.1"
                     /db_xref="GI:13430686"
                     /translation="MLREMLGCCKVYISEARNKTALEAIERALKPFPPAAIVNKFEDA
                     AYGRVGYTVVSSLANGSSSSLKNAVFAMVKTALDTINLELHCGSHPRLGVVDHICFHP
                     LSQTSIEQVSSVANSLAMDIGSILRVPTYLYGAAEKEQCTLDSIRRKLGYFKANREGH
                     EWAGGFDLEMVPLKPDAGPQEVSKAKGVVAVGACGWVSNYNVPVMSNDLKAVRRIARK
                     TSERGGGLASVQTMALVHGEGVIEVACNLLNPSQVGGDEVQGLIERLGREEGLLVGKG
                     YYTDYTPDQIVERYMDLLNNS"
     3'UTR           1727. .1904
                     /gene="At2g20830"
     misc_difference 1742. .1743
                     /gene="At2g20830"
                     /note="compared to genomic sequenc"
                     /replace="atg"
     misc_feature    1889
                     /gene="At2g20830"
                     /note="artifact within poly A tail"
BASE COUNT      523 a    370 c    456 g    555 t
ORIGIN
```

Alignment Scores:

Pred. No.:	0.214	Length:	1904
Score:	111.00	Matches:	31
Percent Similarity:	46.49%	Conservative:	22
Best Local Similarity:	27.19%	Mismatches:	53
Query Match:	14.38%	Indels:	8
DB:	8	Gaps:	2

US-09-980-054A-12 (1-148) x AF360255 (1-1904)

```
Qy           6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
               :::|||  ::::::  |||||||||       :::  ||||||  ||||||
Db         179 CTCGATTGTATCGTGCGTTTAGAGGAGACACATGTTCAACAAGGCTTCGATGAGGGTTAC 238

Qy          26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 :::|||          :::  :::   |||   :::  ||||||
Db         239 GAAGAAGGTCTTGTGTCCGGTCGTGAAGACGCTCGTCATCTGGGTTTGAAACTCGGGTTC 298

Qy          46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------ 63
               :::       :::||||||:::  :::          |||  |||
Db         299 GAGACAGGCGAGCTGATTGGATTCTACAGAGGTTGCTCTGCTCTTTGGAATTCAGCTCTC 358

Qy          64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                 ||||||          |||     |||     |||||||||:::
Db         359 CGTATTGATCCTACACGCTTCTCTCCTCAGCTCCATAAGCATCTCAATGATTTCCATGTC 418

Qy          84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
               ::::::  :::|||  :::  :::  :::   |||  :::|||
Db         419 TTGCTCGATAAAATCCCGCTTTTGGATCCCGAGGACGAAGCTAAAGAC----------- 466
```

```
Qy        104 IleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
                  ::::::  ||||||||       ||||||
Db        467 ------GGGATCAAGGATGATCTCAGAGTCAAATTCAGCATC 502

RESULT 13
AC006234
LOCUS       AC006234               129667 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F5H14 map mi148, complete
            sequence.
ACCESSION   AC006234
VERSION     AC006234.4  GI:20197676
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 129667)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
            Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129667)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 129667)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598541.
FEATURES             Location/Qualifiers
     source          1. .129667
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="mi148"
                     /clone="F5H14"
     misc_feature    1. .29862
                     /note="overlap with BAC clone F26H11
                     (AC006264:63359. .93220)."
     repeat_region   398. .426
                     /rpt_family="AT_rich"
     repeat_region   complement(1061. .1105)
                     /rpt_family="AT_rich"
     repeat_region   complement(1338. .1474)
                     /rpt_family="AT_rich"
     repeat_region   complement(1487. .1891)
                     /rpt_family="Rf:ATREP1|ATREP1 ATREP1 dispersed repeat from
                     GB|D86608 - a consensus."
     repeat_region   complement(1640. .1689)
                     /rpt_family="AT_rich"
     repeat_region   1760. .1797
                     /rpt_family="AT_rich"
     gene            complement(2893. .3840)
                     /gene="At2g21045"
                     /note="synonym: F5H14.48; contains similarity to
                     senescence associated gene Ntdin from GI:7594903
                     (Nicotiana tabacum); contains C2 domain profile
                     (PS50004;PF00168)"
     mRNA            complement(join(<2893. .3015,3089. .3142,3253. .3323,
                     3412. .3485,3566. .3643,3731. .>3840))
                     /gene="At2g21045"
     CDS             complement(join(2893. .3015,3089. .3142,3253. .3323,
                     3412. .3485,3566. .3643,3731. .3840))
                     /gene="At2g21045"
                     /codon_start=1
                     /product="senescence-associated protein"
                     /protein_id="AAM15209.1"
                     /db_xref="GI:20197696"
                     /translation="MYTYSLLNLSHCRRQTRKKRKTDHTEGFLMEETKPKTVEDVETV
                     DVYTAKGFLSTGHRYLDVRTNEEFAKSHVEEALNIPYMFKTDEGRVINPDFLSQVASV
                     CKKDEHLIVACNAGGRGSRACVDLLNEGYDHVANMGGGYSAWVDAGFAGDKPPEDLKI
                     ACKFRPKEN"
     gene            complement(4388. .5541)
                     /gene="At2g21040"
                     /note="synonym: F5H14.40"
     mRNA            complement(join(<4388. .4658,4742. .4831,4920. .5018,
                     5109. .5156,5264. .>5541))
                     /gene="At2g21040"
     CDS             complement(join(4388. .4658,4742. .4831,4920. .5018,
                     5109. .5156,5264. .5541))
                     /gene="At2g21040"
                     /codon_start=1
                     /product="predicted protein"
                     /protein_id="AAM15215.1"
                     /db_xref="GI:20197703"
                     /translation="MQAVQKAPEGTPAAGGMFVVIVHSAEDVEGKHHTNPYVHIYFKG
                     EERKTKHVKKNKDPKWNEEFSFMLEEPPIHEKMHVKVFSTSSRIVFVLRCASVEAAYV
                     PFEGIKPRGAKLFSDGSACSKGFGDTTSFGDMALLGGMAIMLAHHDEGVLTCHVMVRE
                     SFEEINMAWRAWKLVRQGAGSSGSWFVRRRFVRELVRQEAVRQGAGLSGSRFVKEPVR
                     QGAGSSGVGVWPLKSQGRPSDLRFFLKSTFSSFKFHDHYKIIK"
     gene            6120. .6794
                     /gene="At2g21030"
                     /note="synonym: F5H14.46; predicted by genscan"
     mRNA            join(<6120. .6214,6312. .6441,6549. .>6794)
                     /gene="At2g21030"
     CDS             join(6120. .6214,6312. .6441,6549. .6794)
                     /gene="At2g21030"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM15206.1"
                     /db_xref="GI:20197693"
                     /translation="MVKLTCIILENSHITFASLPTGGNDLKRIRFSREMYNKWQAQRR
                     WGENYDKIVELYNVQRFNRQALQTPARSDDQRDSTYSKMDSARESKDWTPRHNFRPPG
                     SVPHHFYGGSSNYGPGSYHGGPPMDAARTTTSSRDDPPSMSNASEMQAEWIEED"
     gene            7487. .8921
                     /gene="At2g21020"
                     /note="synonym: F5H14.6"
     mRNA            join(<7487. .7651,8023. .8247,8523. .>8921)
                     /gene="At2g21020"
     CDS             join(7487. .7651,8023. .8247,8523. .8921)
                     /gene="At2g21020"
                     /codon_start=1
                     /product="putative major intrinsic (channel) protein"
                     /protein_id="AAM15199.1"
                     /db_xref="GI:20197680"
                     /translation="MFGERVIIVPSMAEISDITTQTQTVVSDIENDRSIDDSQSSVLS
                     GPLVSVSFVQKLIGEFVGTFSMIFAGCSAIVVNETYGKPVTLPGIALVWGLVVTVMIY
                     SIGHVSGAHFNPAVSIAFASSKKFPFNQFHPEHYQVHGLMVCYDRQRNHIVRLYVQFA
                     IKYLREIAKKLPKMRFVFYEWLIRLIKTILYVHHLPISMANVMLHVSWRELYRCTTHH
                     TSSISMYFVFSFHITVYIFILYAISDKKLIYSINTKCFLMLNFI"
     repeat_region   9534. .9619
                     /rpt_family="AT_rich"
     gene            complement(10013. .10754)
                     /gene="At2g21010"
                     /note="synonym: F5H14.47"
     mRNA            complement(join(<10013. .10504,10581. .>10754))
                     /gene="At2g21010"
     CDS             complement(join(10013. .10504,10581. .10754))
                     /gene="At2g21010"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM15200.1"
                     /db_xref="GI:20197683"
                     /translation="MGMINPYVQIELSEDKISSKKTTVKHKNLNPEWNEEFKFSVRDP
                     KTQVLEFNVYGWEKIGKHDKMGMNVLALKELAPDERKAFTLELRKTLDGGEEGQPGKY
                     RGKLEVELLYKPFTEEEMQAVQKAPEGTPVAGGMLVVIVHSAEDVEGKHHTNPYVHIY
                     FKGEERKTKNVKKNKDPKWNEEFSFMLEEPPVHEKLHVEVFSTSSRIGLLHPKVTHLF
                     HSQ"
```

```
     gene            complement(11977. .12165)
                     /gene="At2g21000"
                     /note="synonym: F5H14.39"
     mRNA            complement(<11977. .>12165)
                     /gene="At2g21000"
     CDS             complement(11977. .12165)
                     /gene="At2g21000"
                     /codon_start=1
                     /product="putative retroelement pol polyprotein"
                     /protein_id="AAM15202.1"
                     /db_xref="GI:20197685"
                     /translation="MASPLYLHSSDGPENLITNVQLKGENYEDWAKHVRNASRTKQKL
                     GFIDGTLEKPTKEDDMEQ"
     repeat_region   13255. .13286
                     /rpt_family="AT_rich"
     gene            complement(13685. .17125)
                     /gene="At2g20990"
                     /note="synonym: F5H14.5; supported by full length cDNA:
                     Ceres:39669"
     mRNA            complement(join(<13685. .14000,14085. .14573,14646. .14892,
                     14980. .15056,15368. .15442,15555. .15656,15758. .15876,
                     16028. .16151,16247. .16297,16391. .16489,16761. .>17125))
                     /gene="At2g20990"
     CDS             complement(join(13869. .14000,14085. .14573,14646. .14892,
                     14980. .15056,15368. .15442,15555. .15656,15758. .15876,
                     16028. .16151,16247. .16297,16391. .16489,16761. .16871))
                     /gene="At2g20990"
                     /codon_start=1
                     /product="expressed protein"
                     /protein_id="AAM15203.1"
                     /db_xref="GI:20197686"
                     /translation="MGFFSTILGFCGFGVGISLGLVIGYVLFVYLLPNDVKDPEIRSI
                     ADQDPKAMLRMLPEIPLWVKNPDFDRVDWINRFLEYMWPYLDKAICKTAKNIAKPIIE
                     EQIPKYKIDSVEFETLTLGSLPPTFQGMKVYLTDEKELIMEPCLKWAANPNILVAIKA
                     FGLKATVQVVDLQVFAQPRITLKPLVPSFPCFANIYVSLMEKPHVDFGLKLGGADLMS
                     IPGLYRFVQEQIKDQVANMYLWPKTLVVPILDPAKAFRRPVGIVHVKVVRAVGLRKKD
                     LMGGADPFVKIKLSEDKIPSKKTTVKHKNLNPEWNEEFKFSVRDPQTQVLEFSVYDWE
                     QVGNPEKMGMNVLALKEMVPDEHKAFTLELRKTLDGGEDGQPPDKYRGKLEVELLYKP
                     FTEEEMPKGFEETQAVQKAPEGTPAAGGMLVVIVHSAEDVEGKHHTNPYVRIYFKGEE
                     RKTKHVKKNRDPRWNEEFTFMLEEPPVREKLHVEVLSTSSRIGLLHPKETLGYVDIPV
                     VDVVNNKRMNQKFHLIDSKNGKIQIELEWRTAS"
     repeat_region   complement(17189. .17240)
                     /rpt_family="AT_rich"
     gene            17842. .19632
                     /gene="At2g20980"
                     /note="synonym: F5H14.42; predicted by genscan and
                     genefinder"
     mRNA            join(<17842. .17932,18124. .18346,18619. .18856,18937. .19061,
                     19392. .>19632)
                     /gene="At2g20980"
     CDS             join(17842. .17932,18124. .18346,18619. .18856,18937. .19061,
                     19392. .19632)
                     /gene="At2g20980"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAM15204.1"
                     /db_xref="GI:20197687"
                     /translation="MENDQEDLELLLSLDDRVLETPPGSPSAAPVGLFLSLMRGNVCA
                     GYLTDDESPKRRGHSDLSDFRSVVQDCIDYNPKPIAKNTKPKGSNNSNTNDIEKFSGL
                     RIRNLLMGDKLSGCWATMGVITEKGQPKTSSIGQAYGIWKIGSLNENNVSLFLFGDAY
                     KKNETEKAGTVFGLFNCSLRKDNGGREFSLSVNSAKQMVKLGVSADYGVCTAKRKDGT

Alignment Scores:
Pred. No.:              24.4        Length:        129667
Score:                  111.00      Matches:       31
Percent Similarity:     46.49%      Conservative:  22
Best Local Similarity:  27.19%      Mismatches:    53
Query Match:            14.38%      Indels:        8
DB:                     8           Gaps:          2

US-09-980-054A-12 (1-148) x AC006234 (1-129667)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
                :::|||   ::::::   |||||||||           :::    ||||||   ||||||
Db      61246 CTCGATTGTATCGTGCGTTTAGAGGAGACACATGTTCAACAAGGCTTCGATGAGGGTTAC 61305

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 :::|||                    :::   :::    |||   :::   ||||||
Db      61306 GAAGAAGGTCTTGTGTCCGGTCGTGAAGACGCTCGTCATCTGGGTTTGAAACTCGGGTTC 61365

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------ 63
                :::        :::||||||:::   :::                |||   |||
Db      61366 GAGACAGGCGAGCTGATTGGATTCTACAGAGGTTGCTCTGCTCTTTGGAATTCAGCTCTC 61425

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                 ||||||               |||      |||      |||||||||:::
Db      61426 CGTATTGATCCTACACGCTTCTCTCCTCAGCTCCATAAGCATCTCAATGATTTCCATGTC 61485

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              ::::::   :::|||   :::   :::   :::      |||   :::|||
Db      61486 TTGCTCGATAAAATCCCGCTTTTGGATCCCGAGGACGAAGCTAAAGAC------------ 61533

Qy        104 IleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
                          ::::::|||||||||      ||||||
Db      61534 ------GGGATCAAGGATGATCTCAGAGTCAAATTCAGCATC 61569

RESULT 14
AC137683/c
LOCUS       AC137683               58350 bp    DNA     linear   HTG 27-NOV-2002
DEFINITION  Homo sapiens chromosome 11 clone CTD-2063A20 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC137683
VERSION     AC137683.1  GI:25707487
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 58350)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone CTD-2063A20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58350)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28645
                Center clone name: 2063_A_20
            --------------.
```

```
* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      713: contig of 713 bp in length
*      714      813: gap of 100 bp
*      814     1529: contig of 716 bp in length
*     1530     1629: gap of 100 bp
*     1630     2354: contig of 725 bp in length
*     2355     2454: gap of 100 bp
*     2455     3192: contig of 738 bp in length
*     3193     3292: gap of 100 bp
*     3293     3991: contig of 699 bp in length
*     3992     4091: gap of 100 bp
*     4092     4823: contig of 732 bp in length
*     4824     4923: gap of 100 bp
*     4924     5649: contig of 726 bp in length
*     5650     5749: gap of 100 bp
*     5750     6467: contig of 718 bp in length
*     6468     6567: gap of 100 bp
*     6568     7279: contig of 712 bp in length
*     7280     7379: gap of 100 bp
*     7380     8056: contig of 677 bp in length
*     8057     8156: gap of 100 bp
*     8157     8881: contig of 725 bp in length
*     8882     8981: gap of 100 bp
*     8982     9701: contig of 720 bp in length
*     9702     9801: gap of 100 bp
*     9802    10517: contig of 716 bp in length
*    10518    10617: gap of 100 bp
*    10618    11348: contig of 731 bp in length
*    11349    11448: gap of 100 bp
*    11449    12185: contig of 737 bp in length
*    12186    12285: gap of 100 bp
*    12286    13001: contig of 716 bp in length
*    13002    13101: gap of 100 bp
*    13102    13812: contig of 711 bp in length
*    13813    13912: gap of 100 bp
*    13913    14639: contig of 727 bp in length
*    14640    14739: gap of 100 bp
*    14740    15450: contig of 711 bp in length
*    15451    15550: gap of 100 bp
*    15551    16279: contig of 729 bp in length
*    16280    16379: gap of 100 bp
*    16380    17114: contig of 735 bp in length
*    17115    17214: gap of 100 bp
*    17215    17949: contig of 735 bp in length
*    17950    18049: gap of 100 bp
*    18050    18769: contig of 720 bp in length
*    18770    18869: gap of 100 bp
*    18870    19604: contig of 735 bp in length
*    19605    19704: gap of 100 bp
*    19705    20394: contig of 690 bp in length
*    20395    20494: gap of 100 bp
*    20495    21232: contig of 738 bp in length
*    21233    21332: gap of 100 bp
*    21333    22059: contig of 727 bp in length
*    22060    22159: gap of 100 bp
*    22160    22877: contig of 718 bp in length
*    22878    22977: gap of 100 bp
*    22978    23707: contig of 730 bp in length
*    23708    23807: gap of 100 bp
*    23808    24531: contig of 724 bp in length
*    24532    24631: gap of 100 bp
*    24632    25363: contig of 732 bp in length
*    25364    25463: gap of 100 bp
*    25464    26197: contig of 734 bp in length
*    26198    26297: gap of 100 bp
*    26298    27036: contig of 739 bp in length
*    27037    27136: gap of 100 bp
*    27137    27880: contig of 744 bp in length
*    27881    27980: gap of 100 bp
*    27981    28707: contig of 727 bp in length
*    28708    28807: gap of 100 bp
*    28808    29527: contig of 720 bp in length
*    29528    29627: gap of 100 bp
*    29628    30351: contig of 724 bp in length
*    30352    30451: gap of 100 bp
*    30452    31135: contig of 684 bp in length
*    31136    31235: gap of 100 bp
*    31236    31953: contig of 718 bp in length
*    31954    32053: gap of 100 bp
*    32054    32774: contig of 721 bp in length
*    32775    32874: gap of 100 bp
*    32875    33606: contig of 732 bp in length
*    33607    33706: gap of 100 bp
*    33707    34436: contig of 730 bp in length
*    34437    34536: gap of 100 bp
*    34537    35273: contig of 737 bp in length
*    35274    35373: gap of 100 bp
*    35374    36104: contig of 731 bp in length
*    36105    36204: gap of 100 bp
*    36205    36926: contig of 722 bp in length
*    36927    37026: gap of 100 bp
*    37027    37741: contig of 715 bp in length
*    37742    37841: gap of 100 bp
*    37842    38549: contig of 708 bp in length
*    38550    38649: gap of 100 bp
*    38650    39370: contig of 721 bp in length
*    39371    39470: gap of 100 bp
*    39471    40208: contig of 738 bp in length
*    40209    40308: gap of 100 bp
*    40309    41033: contig of 725 bp in length
*    41034    41133: gap of 100 bp
*    41134    41856: contig of 723 bp in length
*    41857    41956: gap of 100 bp
*    41957    42701: contig of 745 bp in length
*    42702    42801: gap of 100 bp
*    42802    43537: contig of 736 bp in length
*    43538    43637: gap of 100 bp
*    43638    44362: contig of 725 bp in length
*    44363    44462: gap of 100 bp
*    44463    45182: contig of 720 bp in length
*    45183    45282: gap of 100 bp
*    45283    46002: contig of 720 bp in length
*    46003    46102: gap of 100 bp
*    46103    46814: contig of 712 bp in length
*    46815    46914: gap of 100 bp
*    46915    47643: contig of 729 bp in length
*    47644    47743: gap of 100 bp
*    47744    48473: contig of 730 bp in length
*    48474    48573: gap of 100 bp
*    48574    49292: contig of 719 bp in length
*    49293    49392: gap of 100 bp
*    49393    50130: contig of 738 bp in length
*    50131    50230: gap of 100 bp
*    50231    50961: contig of 731 bp in length
*    50962    51061: gap of 100 bp
*    51062    51791: contig of 730 bp in length
*    51792    51891: gap of 100 bp
*    51892    52609: contig of 718 bp in length
*    52610    52709: gap of 100 bp
*    52710    53414: contig of 705 bp in length
*    53415    53514: gap of 100 bp
*    53515    54247: contig of 733 bp in length
*    54248    54347: gap of 100 bp
*    54348    55080: contig of 733 bp in length
*    55081    55180: gap of 100 bp
*    55181    55889: contig of 709 bp in length
```

```
            *    55890    55989: gap of 100 bp
            *    55990    56695: contig of 706 bp in length
            *    56696    56795: gap of 100 bp
            *    56796    57517: contig of 722 bp in length
            *    57518    57617: gap of 100 bp
            *    57618    58350: contig of 733 bp in length.

Alignment Scores:
Pred. No.:                19.2          Length:         58350
Score:                    108.00        Matches:        39
Percent Similarity:       51.80%        Conservative:   33
Best Local Similarity:    28.06%        Mismatches:     41
Query Match:              13.99%        Indels:         26
DB:                       2             Gaps:           7

US-09-980-054A-12 (1-148) x AC137683 (1-58350)

Qy         22 LysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyr 41
              :::||||||:::::|||      |||:::|||   :::      |||:::   |||
Db       9684 GAGGAAGGTGAAGTCCAATTTTAAAAAAACCAAAANATATGATACAAGAAATGAGGGGAG 9625

Qy         42 GlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrp 61
              ::::::      :::             |||      ||||||          |||   |||
Db       9624 AAATCTTCAGTTAAAGAGATAGCATAAATAAAAAACATTCAAAACTTCAGGAAACAATGG 9565

Qy         62 LeuSerHisIleAspGlnTyrAsnAsnSerSer--------------------------- 72
              ::::::           |||   |||:::
Db       9564 ATGAATGGATGC---ACTTATAGAAATGCAAAATGCTCTAGAAAGTCTCAGCAATAGAAT 9508

Qy         73 ---SerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThr 91
                 |||   ||||||      :::   |||||||||   :::   :::::::::   :::
Db       9507 CCAAGCAGAAGAAAT------TCAGAGCTTGAAGACAAAGTTTTTGAATTAACCCAATCC 9454

Qy         92 AsnGlyAspLysGlu------ValGluAspTyrGluLysAsnIleLysLys--------- 106
              |||   |||||||||      :::      |||||:::::::::::::::
Db       9453 AACAAAGACAAAGAAAAAAGAATAAGAAAATATGAACAAAGCCTCCAAGAAATCTGGGAT 9394

Qy        107 ------AlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIle 124
                  ::::::   ||||||:::|||   ||||||   ::::::      |||
Db       9393 TATGTTAAACAATCAAACCTAAGAATAATC---AGCATTCCTGAGAAAGAAGAGAAATCT 9337

Qy        125 AspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnValSerGluAsn 143
                |||||||:::||||||   :::|||:::               :::   ||||||
Db       9336 AAAAGTTTGGAAAACTTATCCAGGGAAATA--------------ATCAAGGAAAAC 9295

RESULT 15
AE001581
LOCUS       AE001581               30651 bp    DNA     circular BCT 17-MAR-2003
DEFINITION  Borrelia burgdorferi B31 plasmid cp32-9, complete plasmid sequence.
ACCESSION   AE001581
VERSION     AE001581.1  GI:6382346
KEYWORDS    .
SOURCE      Borrelia burgdorferi B31
  ORGANISM  Borrelia burgdorferi B31
            Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
            Borrelia burgdorferi group.
REFERENCE   1  (bases 1 to 30651)
  AUTHORS   Casjens,S., Palmer,N., van Vugt,R., Huang,W.M., Stevenson,B.,
            Rosa,P., Lathigra,R., Sutton,G., Peterson,J., Dodson,R.J., Haft,D.,
            Hickey,E., Gwinn,M., White,O. and Fraser,C.M.
  TITLE     A bacterial genome in flux: the twelve linear and nine circular
            extrachromosomal DNAs in an infectious isolate of the Lyme disease
            spirochete Borrelia burgdorferi
  JOURNAL   Mol. Microbiol. 35 (3), 490-516 (2000)
  MEDLINE   20138354
   PUBMED   10672174
REFERENCE   2  (bases 1 to 30651)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 30651)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
            White,O., Dodson,R., Hickey,E.K., Gwinn,M., Peterson,J.,
            van-Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    Amino acid sequence update by submitter
FEATURES             Location/Qualifiers
     source          1. .30651
                     /organism="Borrelia burgdorferi B31"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:224326"
                     /plasmid="cp32-9"
     gene            66. .1292
                     /gene="BBN01"
     CDS             66. .1292
                     /gene="BBN01"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07681.1"
                     /db_xref="GI:6382369"
                     /translation="MCDLRKTKLIDKISSLELYKYSIFFRNYIENVAEDCLKNGLVLE
                     SADHNVSEVELARLKVQLKNALLNCIISYRFHGIGYVLVKTKDTLIDLEQPVNIELPI
                     GFEYLDYEYVRDLGVDFDHITYKVKSNNKNNSLDAVKIHKSRLIIYENFDYILKRYVP
                     CYTESFLLDIYLPEKIYVEIERRIENHNFLFYKDESLVQLQDALSSATTSLSALTKSS
                     NNDGGSGILSSFLRKQNSNNHSKDISNLRNLNDSLSQELARLKSNLNNEGMFYTATLS
                     ASLEVIKYDLSYLREALALIKAKIGADTKEPLTRSFNEQAKGLGNGGKGDRSNYYDFL
                     KGVQEQIENSCNLKLTKYFGLDMKFNSLIMLSEEQKVERDIKLIELYSKYNQLIQSSS
                     FDNEELAILKEKLFSF"
     gene            1309. .2001
                     /gene="BBN02"
     CDS             1309. .2001
                     /gene="BBN02"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07682.1"
                     /db_xref="GI:6382370"
                     /translation="MTEKEEKEDLQAQDKEEQQNKADTKVISAQEFEEYMRFKEQANS
                     KSKETSRDLSINERITKELAEVEERERIKKQLLLEAERINEIDTLAKAHLSNHFNKEV
                     LLAKGYTLKDIMQAQRRELVRKFVPIEQIKAIAKVSDISHIDGEMLEQLVSLAKVNIK
                     LRKTASSNFSSVDSVRGNIVVKSEERASLLDSNFVPINFTEFVQAISNTYKQRRIQFY
                     ENLKRHKRTSIA"
     gene            2013. .2579
                     /gene="BBN03"
     CDS             2013. .2579
                     /gene="BBN03"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07683.1"
                     /db_xref="GI:6382371"
                     /translation="MSDGITKIKEEFDKKVAEIKALMKNPQQDAGLLRNSVDFRDKNL
                     IYSNSDGVFTSSKDKIENYPAKGYPYKRGVKLSFSADGTTELEVEAGGGDDLYGICTD
                     IDEFTGMATVVPIINNFTGYLTFKKNGQNGVNPGDKLHFNAQGELEKNGGNDKSVNAI
                     ALSKVHKLTEELSIVLASVFGNRALKGN"
     gene            2583. .3350
                     /gene="BBN04"
     CDS             2583. .3350
                     /gene="BBN04"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
```

```
                     /protein_id="AAF07684.1"
                     /db_xref="GI:6382372"
                     /translation="MALKGKGQAKSPNVHDNPQLGLESEIPVAPRSKRQARQAEEAQA
                     KDPYLDSVKELDDVLLKFKKYSKSMSSIENKVFSSSGGCFKSKNERVNAYSFTYSSFA
                     DKIEEYLYDPANSFPYKRGVKLVPKENSIYVEVGADTDMYGICVDVCEFSSTAYVLPI
                     TNNFEGYLVTRNPSIKMGEILDINNNGVIIKAGGGPPTAINAYALSDSFTINFAPEDE
                     HQDQARYPKQEYSINLIKVAIFGNRGLEKTVTPEAGG"
gene                 4958. .5413
                     /gene="BBN07"
CDS                  4958. .5413
                     /gene="BBN07"
                     /note="similar to GP:1663562 percent identity: 98.68;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF07659.1"
                     /db_xref="GI:6382347"
                     /translation="MSEQKNLQTQVEAEEELLVTKLHSEVLLLLGIDKLALSRQNFLL
                     HLSLLQAILVTRGIDASSLTYEQIFLLTFYHMGCQLRKQGVVREFEFDRIKKEKFNEL
                     ELDYYPSSSGGEEGGEGGCGSNKSFCSQPDAFLEKLKRETSTPSCVGVV"
gene                 5395. .5796
                     /gene="BBN08"
CDS                  5395. .5796
                     /gene="BBN08"
                     /note="similar to GP:1663563 percent identity: 99.21;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF07660.1"
                     /db_xref="GI:6382348"
                     /translation="MCGGCLMNGVRKRLSDMSFRMINVFKDPKPLKFYKGTVVKLEND
                     SSYQRVFDKNKYTEFAGVIIDIKPQELAILYDSDMSDIQGYSKLYTYQDLNYELKDRI
                     SIADLVYFEIFSIDSSIGYFTLVLKEFIWTN"
gene                 5784. .6173
                     /gene="BBN09"
CDS                  5784. .6173
                     /gene="BBN09"
                     /note="similar to GP:1663564 percent identity: 100.00;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF07661.1"
                     /db_xref="GI:6382349"
                     /translation="MDKLEFKMELEIGWPGGRAGIAKMHEKGSSNLPARKHLTKIASS
                     SEFREYINNSYINSKFNLDPKSGMEAIGQAFIRYYENYLLSAQVTPALKANTIKSKFK
                     RGSNTAAIPLVDTAKMLSEITYKVTLE"
gene                 6170. .6736
                     /gene="BBN10"
CDS                  6170. .6736
                     /gene="BBN10"
                     /note="similar to GP:1663565 percent identity: 97.98;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF07662.1"
                     /db_xref="GI:6382350"
                     /translation="MIFTLDMVLNHLTQIFKGFKAYATENNFECDIINTYNHPYLSKI
                     TAASSNIIALKFDGTENLFDHNYKAGVFYENALEFSINFQIYIIAIVLNAKDFDANSR
                     MLMLYSMLSDFLHNKAHKYTLPSLQPEYINKINFYIYPTSNMQTVGLINLGTKYSNHA
                     YSASIAFNASVKAIEILKEEYEIAARYN"
gene                 6717. .7805
                     /gene="BBN11"
CDS                  6717. .7805
                     /gene="BBN11"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07685.1"
                     /db_xref="GI:6382373"
                     /translation="MPQDTISVSFIDSRIQTSRPNYYNPLLVYKTAKIKVNKDAASYK
                     ILNLTVNNYEKQIETLEKENGNEEDQFGKEKTLLKTAMSNFFNSSEESLKSADLFIYK
                     DKPEELKNYLKVHRHTFVVLINTEGDASDDGLKIYKDDYNKFKMPSTFFVFSTKEQEI
                     KELFKDKGNTEKERNIAVYSNNKDNLHLKFISQYLHQASIFHAVNPYSMPLAATPLVD
                     DTVIGKLRTAKINFYSLLNETGLDGLPAFKEGVDLAGSTIDEQFTYHYIKNEAIIELI
                     KIWNKNNRQNSKLSALQLSGARDNAYTSAIECLLKRFVDRGLIIEYKNLNLTLSPTPQ
                     LKLELGVNITYNFSINAVALVITTQDIV"
gene                 7845. .8273
                     /gene="BBN12"
CDS                  7845. .8273
                     /gene="BBN12"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF07686.1"
                     /db_xref="GI:6382374"
                     /translation="MQFYDLREVYFSIGGTQLHSGKLELTSEPTTRAVISSEDKGMPV
                     ISLRDPKTITYVFNIEVTLGSHDYILLTELSDEQFYNMDVRKEDKMLDLVFNDRIATK
                     IISNYAIFTEEPSRSYSAEAEKVSFEIRAINCQKTKPNNT"
gene                 8742. .8978
                     /gene="BBN14"
CDS                  8742. .8978
                     /gene="BBN14"
                     /note="hypothetical protein; identified by Glimmer2;
```

```
Alignment Scores:
Pred. No.:                10.4          Length:          30651
Score:                    107.50        Matches:         46
Percent Similarity:       43.21%        Conservative:    24
Best Local Similarity:    28.40%        Mismatches:      55
Query Match:              13.92%        Indels:          37
DB:                       1             Gaps:            9

US-09-980-054A-12 (1-148) x AE001581 (1-30651)

Qy         13 GluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGly---ThrLysAsp 31
              ::::::|||   |||:::          |||   |||:::      |||        :::
Db      22043 AAGCAAGAAAGTTATGATTTTTATAAAAAAAATGGGAAGTTTACTGGGTTTTTATTGGAA 22102

Qy         32 GlnTyrLeuGluGly----------------------------LysGluTyr--------- 39
              :::  ||||||                              |||:::|||
Db      22103 GAACTTCTTGAAAGTCAAACAGATTTGATTAATAAGCTTTTAAAAAAATATAAACAATTA 22162

Qy         40 ---GlyTyrGlnThrGlyPheGlnArg-------------------------------- 47
                |||||      |||   |||
Db      22163 AAAGGATATTAAGAGGATTTTATGAGAAATTTGGTGCACAGAACATATGATATAGAAAGC 22222

Qy         48 ------------PheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
                         ||||||  ||||||:::   :::|||   :::   |||
Db      22223 ATAAAAAATGAATTTTTAAACATAGGATTTAGTGAAGAGGCAATAGATTTTGTTTTTCTT 22282

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
              |||   |||   |||   |||          |||:::      :::   ::::::|||
Db      22283 CATAATGATAATTAC---AACTATGAGGTTTTAAAAGAGAAAATAATTGATGTAGAAAAG 22339

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              :::   |||     |||::::::  |||   ::::::::::::   ||||||||||
Db      22340 AATTTG---CAAAAAGACATATCTAGTTTAGATACTAAGATAGATAATGTAGAAAAGAAT 22396

Qy        104 ------------IleLysLysAlaArgAsnLysLeuArgVal---IleAlaSerIleThr 118
                         |||                |||   :::   |||   |||:::
Db      22397 TTAAATGTTAAGATAGATAATGTAGAGAAGAATTTAAATATTAAAATAGATAGTGTTAAA 22456

Qy        119 LysGlu---ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThr 137
                |||       ||||||||||||||||||      :::      |||      |||
Db      22457 AATGAACTTAATTCTAAAATAGATAGTTTAGATACTAAGATAGATAATGTAGAAAAAACT 22516

Qy        138 LeuGln 139
                ||||||
```

Db 22517 TTGCAA 22522

Search completed: February 16, 2004, 15:27:47
Job time : 3247 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:12:22 ; Search time 20 Seconds
(without alignments)
711.648 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	223	28.9	198	2	S63233	hypothetical prote
2	132	17.1	134	2	T41221	hypothetical prote
3	116	15.0	404	2	H84593	hypothetical prote
4	97.5	12.6	610	2	C70126	DNA mismatch repai
5	94	12.2	627	2	E70122	flagellar hook-ass
6	92	11.9	270	2	C82114	probable flagellar
7	91.5	11.9	1005	2	A64465	hypothetical prote
8	91	11.8	239	2	AD0088	probable flagellar
9	91	11.8	2325	2	T15566	hypothetical prote
10	90.5	11.7	746	2	T47237	myosin II heavy ch
11	87.5	11.3	755	2	T34567	hypothetical prote
12	87	11.3	263	2	E70215	hypothetical prote
13	86.5	11.2	212	2	AH2423	phosphoglycolate p
14	86	11.1	2285	2	T12796	probable transglyc
15	85	11.0	507	2	A71622	hypothetical prote
16	85	11.0	805	2	A75014	hypothetical prote
17	85	11.0	1553	2	S67483	adenosinetriphosph
18	84.5	10.9	852	2	D72230	conserved hypothet
19	84.5	10.9	1188	2	E89896	chromosome segrega
20	84	10.9	1269	2	F84730	probable myosin he
21	83.5	10.8	592	2	C70302	conserved hypothet
22	83.5	10.8	720	2	G97229	pullulanase [impor
23	83.5	10.8	879	2	C71083	conserved hypothet
24	83.5	10.8	880	2	F75103	conserved hypothet
25	83.5	10.8	909	2	C97325	hypothetical prote
26	83.5	10.8	1231	2	S70553	chromosome-associa
27	83	10.8	777	2	C85065	kinesin-like prote
28	82.5	10.7	244	2	H71873	hypothetical prote
29	82.5	10.7	981	2	G90601	DNA polymerase III
30	82.5	10.7	1225	2	A49464	chromosome segrega
31	82	10.6	238	2	T19175	hypothetical prote
32	82	10.6	1164	2	T24806	hypothetical prote
33	82	10.6	1516	2	E71619	RAD2 endonuclease
34	82	10.6	1963	1	MWKW	myosin heavy chain
35	81	10.5	320	2	AH0942	6-phosphofructokin
36	81	10.5	500	2	E89852	hypothetical prote
37	81	10.5	521	2	B70182	lysine-tRNA ligase
38	81	10.5	564	2	A60115	M protein precurso
39	81	10.5	797	2	E84642	hypothetical prote
40	81	10.5	1188	2	A71621	protein with 5'-3'
41	81	10.5	2429	1	SJHUA	spectrin alpha cha
42	81	10.5	3394	2	T18501	hypothetical prote
43	80.5	10.4	211	2	S69930	outer surface prot
44	80.5	10.4	263	2	B96957	probable transcrip
45	80.5	10.4	402	2	T23701	hypothetical prote

ALIGNMENTS

RESULT 1
S63233
hypothetical protein YNL260c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N0838
C;Species: Saccharomyces cerevisiae
C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C;Accession: S63233
R;Sen-Gupta, M.; Gueldener, U.; Beinhauer, J.; Fiedler, T.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63220
A;Accession: S63233
A;Molecule type: DNA
A;Residues: 1-198 <SEN>
A;Cross-references: EMBL:Z71536; NID:g1302312; PID:e239647; PID:g1302313; GSPDB:GN00014;
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YNL260c
A;Cross-references: SGD:S0005204
A;Map position: 14L

Query Match 28.9%; Score 223; DB 2; Length 198;
Best Local Similarity 35.9%; Pred. No. 1.4e-10;
Matches 47; Conservative 30; Mismatches 40; Indels 14; Gaps 3;

```
Qy          4 IDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS 63
              :| ||:|||||: |: || ||| :  |  :||||:|| | ||||| ::| :: |
Db         37 MDFDNLLNLEEQYYQEGFLEGQNENIKQSFLEGKQYGLQVGFQRFTLLGQMEGLCDV--- 93

Qy         64 HIDQYN-NSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
                |: |  :| :|  ::: :  :|  : : | |: | ::|: : | :|| | |
Db         94 -IESYGLHSPTLEKNIHTIRTLMKGLKMNNDDESVMEFERVLIKLKNKFRTI-------- 144

Qy        123 KIDSLDNLVKE 133
                : :|  |||:
Db        145 -LITLHRLVKD 154
```

RESULT 2
T41221
hypothetical protein SPCC191.08 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41221
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z21904
A;Accession: T41221
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-134 <LYN>
A;Cross-references: EMBL:AL049644; PIDN:CAB41054.1; GSPDB:GN00066; SPDB:SPCC191.08
A;Experimental source: strain 972h-; cosmid c191

```
C;Genetics:
A;Gene: SPDB:SPCC191.08
A;Map position: 1

  Query Match             17.1%;  Score 132;  DB 2;  Length 134;
  Best Local Similarity   28.0%;  Pred. No. 0.0013;
  Matches   37;  Conservative   29;  Mismatches   58;  Indels    8;  Gaps    5;

Qy          4 IDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS 63
              :| : |  ||| :|: |: || ::| :  | |   :| :  : ::|: | |    : |||
Db          1 MDWEEVTKLEENEYKRGYDEGILKGIEQGYEEAFLFGLEHAYNKYLLAGEIYGRVCFWLK 60

Qy         64 HID-QYNNSSSLRNHLNNLEDIMAQISITNGDKEVED--YEKNIKKARNKLRVIASI--T 118
                : |:         ||  |: ::  :  || : |  |  ::     |   | :|::|:  |
Db         61 EENSQHPKIKKAHRHLEQLKSLLESLP-TNNELEETDAGFDSYWNKITAKAKVVSSLLGT 119

Qy        119 K--ETWKIDSLD 128
               |     |||: |
Db        120 KILPAEKIDAND 131


RESULT 3
H84593
hypothetical protein At2g20830 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84593
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84593
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: GB:AE002093; NID:g4454465; PIDN:AAD20912.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g20830
A;Map position: 2

  Query Match             15.0%;  Score 116;  DB 2;  Length 404;
  Best Local Similarity   26.0%;  Pred. No. 0.082;
  Matches   32;  Conservative   22;  Mismatches   51;  Indels   18;  Gaps    2;

Qy          6 IDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS-- 63
              :| :: ||| : || ||  :|      : :  | : ||:   :||: :      | |
Db         10 LDCIVRLEETHVQQGFDEGYEEGLVSGREDARHLGLKLGFETGELIGFYRGCSALWNSAL 69

Qy         64 HIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKN----------------IKKA 107
                ||    |  | |||:   :: :| : : : | :|   |:                | :|
Db         70 RIDPTRFSPQLHKHLNDFHVLLDKIPLLDPEDEAKDGIKDDLREMLREMLGCCKVYISEA 129

Qy        108 RNK 110
               |||
Db        130 RNK 132


RESULT 4
C70126
DNA mismatch repair protein (mutL) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 17-Mar-2000
C;Accession: C70126
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: C70126
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-610 <KLE>
A;Cross-references: GB:AE001131; GB:AE000783; NID:g2688098; PIDN:AAC66594.1; PID:g268805
A;Experimental source: strain B31
C;Superfamily: mismatch repair protein hexB

  Query Match             12.6%;  Score 97.5;  DB 2;  Length 610;
  Best Local Similarity   24.7%;  Pred. No. 3.8;
  Matches   37;  Conservative   27;  Mismatches   51;  Indels   35;  Gaps    6;

Qy          1 MSDIDIDNVLNLEE-----EQYELGFKE-GQIQGTKDQYLEGKEYGYQTGFQRFLIIGYI 54
              : :|  || |  ||     |: |: |   |||                  |  |||:  |
Db        407 IGNIFSDNFLEFEEPPNKNEKEEIKFNYIGQI-----------------FSEFLIVEKI 448

Qy         55 QELMKFWLSHIDQYNNS-----SSLRNHLNNLEDIMAQISITNGDKEVED-YEKNIKKAR 108
                |:       |||:          |||   |:: ::  |  |  || :|:  :  |:: :
Db        449 NEIY-----FIDQHAVHEKIIYEKLRNSKKNVQKLLVPIEFTVVDKNIEEIIDSEIEEYK 503

Qy        109 NKLRVIASITKETWKIDSLDNLVKEVGGTL 138
                :|: |  : ::::|: |:  :   ||
Db        504 QMDIIISKIGPKKYQLESIPNICSQYENTL 533


RESULT 5
E70122
flagellar hook-associated protein (flgK) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: E70122
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70122
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-627 <KLE>
A;Cross-references: GB:AE001129; GB:AE000783; NID:g2688071; PIDN:AAC66579.1; PID:g268804
A;Experimental source: strain B31

  Query Match             12.2%;  Score 94;  DB 2;  Length 627;
  Best Local Similarity   25.7%;  Pred. No. 7.3;
  Matches   49;  Conservative   23;  Mismatches   51;  Indels   68;  Gaps   10;

Qy         10 LNLEEEQYELGFKEG----QIQGTKDQYL------EGKEYGYQTGFQRFLII-------- 51
              ||  ::| :||  :|      |   ||: |      |     || |   :|: |
Db         55 LNRAKKQGQLG--QGIVVQSIDRVKDELLNTRIIEESHRLGYWTSQDKFISILEDVYNEP 112

Qy         52 ---GYIQELMKFWLSHIDQYNNSSSL----------------RNHLNNLED--IMA--Q 87
                  : |  || |  |  |     |                  ||  ::||   |||  :
Db        113 EDQSIRKRLNDFWESWHDLANQPQGLAERKIILERGKSFCEGIRNRFHSLERIYIMANDE 172

Qy         88 ISITNGDKEVEDYEKNIKKARNKLRVIASITKETWK----------IDSLDNLVKEVGG 136
              | ||  :          | | :| ||:: |:  |           :|: | :|:::|
Db        173 IKITTDE-----------ANNYIRNIANLNKQISKSQAMKDNPNDLMDARDLMVEKLGN 220

Qy        137 TLQVS-ENPDD 146
               : || ||  |
Db        221 IISVSIENKQD 231


RESULT 6
C82114
probable flagellar assembly protein FliH VC2131 [imported] - Vibrio cholerae (strain N1
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: C82114
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
```

```
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: C82114
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-270 <HEI>
A;Cross-references: GB:AE004286; GB:AE003852; NID:g9656679; PIDN:AAF95276.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2131
A;Map position: 1

  Query Match             11.9%;  Score 92;  DB 2;  Length 270;
  Best Local Similarity   22.3%;  Pred. No. 4;
  Matches   33;  Conservative   31;  Mismatches   42;  Indels   42;  Gaps    6;

Qy         13 EEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSS 72
              : | |: ||::|: :| :  : ||:  ||| |      :  ||| :| :::  :|:
Db         88 QAEGYQQGFEQGKAEGFQAGHQEGQTQGYQDGVAEGQAL--IQEQVKTFMALANQFAQPL 145

Qy         73 SLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKID------- 125
               | |          ||:     :|:: |              :: ::|||   ::
Db        146 DLLN---------AQV-----EKQLVD-------------MVLALTKEVVHVEVQTNPQV 178

Qy        126 SLDNLVKEV------GGTLQVSENPDDM 147
                ||  :   |      |  : :  ||:|:
Db        179 ILDTVKASVEALPIAGHAITLKLNPEDV 206


RESULT 7
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: A64465
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: A64465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1005 <BUL>
A;Cross-references: GB:U67572; GB:L77117; NID:g1591958; PIDN:AAB99331.1; PID:g1591962; T
C;Genetics:
A;Map position: REV1273394-1270377
C;Superfamily: hypothetical protein MJ1322

  Query Match             11.9%;  Score 91.5;  DB 2;  Length 1005;
  Best Local Similarity   24.7%;  Pred. No. 20;
  Matches   43;  Conservative   34;  Mismatches   62;  Indels   35;  Gaps    7;

Qy           3 DID-IDNVLN-------------LEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQR 47
               ||| :| :||             | ||   |  :  :|:  |    | |||     :::
Db         342 DIDNLDTLLNKIKDEIERVETIKDLLEELKNLNEEIEKIEKYKRICEECKEY-----YEK 396

Qy          48 FL-IIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI----------SITNGDKE 96
               :| :     |  |  | :|       |:  ::|:||  : ::           || |  ||
Db         397 YLELEEKAVEYNKLTLEYITLLQEKKSIEKNINDLETRINKLLEETKNIDIESIENSLKE 456

Qy          97 VEDYEK---NIKKARNKL-RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDD 146
               :|: :|   |::| : :| : :  |  |  ::  : : :||| |   : : | |
Db         457 IEEKKKVLENLQKEKIELNKKLGEINSEIKRLKKILDELKEVEGKCPLCKTPID 510


RESULT 8
AD0088
probable flagellar assembly protein fliH [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AD0088
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AD0088
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89567.1; PID:g15978800; GSPDB:GN00175
C;Genetics:
A;Gene: fliH

  Query Match             11.8%;  Score 91;  DB 2;  Length 239;
  Best Local Similarity   26.5%;  Pred. No. 4.1;
  Matches   41;  Conservative   25;  Mismatches   57;  Indels   32;  Gaps    7;

Qy           7 DNVLNLEEEQYEL--GFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
               |  |:  | | :|  ||:||  ||       |||| ||| | :    :|:   | |
Db          35 DQTLDPAEYQKQLMAGFQEGISQGFDKGLAEGKEEGYQEGVR----LGHDDGLKK----- 85

Qy          65 IDQYNNSSSLRNHLNNLEDIMAQIS--ITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
                       :: | :  |::    |  ||      :| ||:  ::    |:::  :|::
Db          86 ----GRIEGRQSELASFNDVIKPFSGYITQLHTYLETYEQ--RRRDELLQLVEKVTRQVI 139

Qy         123 KID------SLDNLVKE-------VGGTLQVSENP 144
               : :       |  ||:|       |    |:|  ||
Db         140 RCELALQPAQLLTLVEEALAALPMVPQQLKVYLNP 174


RESULT 9
T15566
hypothetical protein C18H2.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15566
R;Du, Z.
submitted to the EMBL Data Library, April 1994
A;Description: The sequence of C. elegans cosmid C18H2.
A;Reference number: Z18370
A;Accession: T15566
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2325 <DUZ>
A;Cross-references: EMBL:U00040; NID:g474197; PID:g474199; PIDN:AAA50665.1; CESP:C18H2.3
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C18H2.3
A;Introns: 38/1; 476/3; 488/3; 734/2; 792/3; 831/1; 939/3; 1045/1; 1091/3; 1137/1; 1204/

  Query Match             11.8%;  Score 91;  DB 2;  Length 2325;
  Best Local Similarity   22.3%;  Pred. No. 57;
  Matches   33;  Conservative   27;  Mismatches   60;  Indels   28;  Gaps    7;

Qy           4 IDIDNVL---------NLEEEQYELGFK--------EGQIQGT--KDQYLEGKEYGY-QT 43
               ::|: :|          | ::   |  ||          :  ::|   ||:   |:   : :
Db        1415 VEINKILDSRNKTKSSNFDDWVYSTGFSARGDYLKIKNDLKGKWFKDEVSMGRNTDFLEK 1474

Qy          44 GFQRFLIIG-YIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEK 102
                 ::|: :|  |    | ||         |     :  : : :::| :  | |   | |
Db        1475 ALEKFITLGEIISNFEKAWL------NFDKQIEANFSKIDNITSLFSTTESFKNVNSSET 1528

Qy         103 NIKKARNKLRVIASITKE-TWKIDSLDN 129
               ||||  : ||  :   |      | :||
Db        1529 NIKKMEDTLRNCSMFEKTIDGPFDEMDN 1556


RESULT 10
T47237
```

myosin II heavy chain [imported] - Naegleria fowleri (fragment)
C;Species: Naegleria fowleri
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47237
R;Shaw, D.R.; Sullivan, P.K.; Marciano-Cabral, F.; Ennis, H.L.
submitted to the EMBL Data Library, December 1995
A;Description: Codon usage in Naegleria fowleri.
A;Reference number: Z24413
A;Accession: T47237
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-746 <SHA>
A;Cross-references: EMBL:U43192; PIDN:AAB01786.1
A;Experimental source: strain LEE mp; cell type amoeba

Query Match 11.7%; Score 90.5; DB 2; Length 746;
Best Local Similarity 27.7%; Pred. No. 17;
Matches 41; Conservative 28; Mismatches 58; Indels 21; Gaps 5;

```
Qy          5 DIDNVLNLEE---EQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFW 61
              :::|:|:  |     | :  ||: | :   ::    |        || |       | ||
Db        147 ELENLLSDTEGGKNQLDSQFKQLQNELQNERTNLQKMKSENERLQREL------EEMKRS 200

Qy         62 LSHIDQYNNSSSLRNHLNNLEDIMAQI---------SITNGDKEVEDYEKNIKKARNKLR 112
               ||  |: | |:|| : : :||| : ::         | |: ||:    :| :|:  :|:
Db        201 LS--DKQNESTSLDSKVKSLEDKIRELTALLETERSSKTDLDKKRSKMDKEVKRLAQQLQ 258

Qy        113 VIASITK-ETWKIDSLDNLVKEVGGTLQ 139
                  | || | :  || ||::   ||
Db        259 ETEQALKGETQKKNDADNRVKQLESELQ 286
```

RESULT 11
T34567
hypothetical protein DKFZp434A128.1 - human
C;Species: Homo sapiens (man)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34567
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21542
A;Accession: T34567
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-755 <OTT>
A;Cross-references: EMBL:AL122120
A;Experimental source: adult testis; clone DKFZp434A128
C;Genetics:
A;Note: DKFZp434A128.1

Query Match 11.3%; Score 87.5; DB 2; Length 755;
Best Local Similarity 23.7%; Pred. No. 29;
Matches 36; Conservative 39; Mismatches 50; Indels 27; Gaps 8;

```
Qy          3 DIDIDN-VLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFW 61
              | |||| | |   : |   ||  ::  | ::|| | | | |:: : :   ::|:|
Db        322 DGDIDNCALELARIKQETREKENLVK-EKIKFLE-SEIGNNTEFEKRISVA-DRKLLKCR 378

Qy         62 LSHIDQYNNSSSLRNHLNNLEDIMAQISITNGD-----------KEVEDYEKNIKKARN 109
               ::  |   :   |:  |::|:   | :: |: |           |:: :    ::| :|
Db        379 TAYQDHETSRIQLKGELDSLK---ATVNRTSSDLEALRKNISKIKKDIHEETARLQKTKN 435

Qy        110 KLRVIAS----ITKETWKID----SLDNLVKE 133
                :| :    ||::|  ::    :|::::||
Db        436 HNEIIQTKLKEITEKTMSVEEKATNLEDMLKE 467
```

RESULT 12
E70215
hypothetical protein BBA69 - Lyme disease spirochete plasmid A/lp54
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
C;Accession: E70215
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70215
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-263 <KLE>
A;Cross-references: GB:AE000790; NID:g2690224; PIDN:AAC66287.1; PID:g2690286; TIGR:BBA6
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid
C;Superfamily: Borrelia burgdorferi hypothetical protein BBI38

Query Match 11.3%; Score 87; DB 2; Length 263;
Best Local Similarity 23.5%; Pred. No. 9.6;
Matches 42; Conservative 32; Mismatches 47; Indels 58; Gaps 8;

```
Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQYLE--GKE-------YGYQ----TGFQRFLIIGY 53
              :|: | :::  :||  : :  ||    |: |||       |  |    |  :  |:  |
Db         55 ENIQNFKDKSGDLGASDEKFMGTTASELKAIGKELEDRKNQYDIQIAKITNEESNLLDTY 114

Qy         54 IQE----------LMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEK 102
              |:           | :| || :|       : ::  |::|:        :| : :||
Db        115 IRAYELANENEKMLLKRFLLSSLDY------KKENIETLKEIL--------EKLINNYEN 160

Qy        103 NIKKARNKLRVIA------------------SITKETWKIDSLDNLVKEVGGTLQVSE 142
              : | | | |  ||                  :::||  | | |: |:::|   ||: |
Db        161 DPKIAANFLYRIALDIQLKLEKHLKSINEKLDTLSKENSKED-LEALLEQVKSALQLQE 218
```

RESULT 13
AH2423
phosphoglycolate phosphatase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2423
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2423
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-212 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76643.1; PID:g17134082; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4944
C;Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 11.2%; Score 86.5; DB 2; Length 212;
Best Local Similarity 29.5%; Pred. No. 8.2;
Matches 38; Conservative 16; Mismatches 48; Indels 27; Gaps 7;

```
Qy         26 IQGTKDQYLEGKEYGYQTGF---------QRFLIIGYIQELMKFWLSHIDQYNNSSSLRN 76
              | | |:  || ||: |: |           || |  :  |  |  | :  :  :: : |
Db         85 IPGIKEALLELKEHDYKLGIITSNSRENVTNFLSINELDSLFDFIYSGVTIFGKTTIINN 144

Qy         77 HLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNLVKEVGG 136
              |   :   |  |  || | | | : |||    ::|||      ||  :| : | |
Db        145 VLRQ-KQFKPQSVIYVGD-ETRDIEAS-KKA--NIKVIA----VTWGFNSPEILAK---- 191

Qy        137 TLQVSENPD 145
                 :|||
Db        192 -----QNPD 195
```

```
RESULT 14
T12796
probable transglycosylase - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C;Accession: T12796; A69911
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 propha
A;Reference number: Z17583
A;Accession: T12796
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2285 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC13005.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2285 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e1183582;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yomI

  Query Match             11.1%;  Score 86;  DB 2;  Length 2285;
  Best Local Similarity   22.4%;  Pred. No. 1.4e+02;
  Matches   35;  Conservative   34;  Mismatches   61;  Indels   26;  Gaps    6;

Qy           5 DIDNVLNL-EEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGY------IQEL 57
               |:|| ::: ||:: :: :   ||:  : |  | |:|  |   |:    |:      | |
Db        1856 DVDNKISMTEEDEDKVKYYSKQIKLIQQQQKEAKKYIKQLEEQKKAAKGFPDIQEQITEE 1915

Qy          58 MKFWLS-----HIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARN-KL 111
               |: |       ::: ||   |:::   :| | :  |           |::  :| |: :|
Db        1916 MQNWKDKQKDFNLELYNTKKSIKDIYKSLADEVVSI----------YKEMYEKMRDIEL 1964

Qy         112 RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDDM 147
                     |::   || :|     |      :: |  | :
Db        1965 EAHQKATQDL--IDEIDKTDDEAKFQKELKERQDSI 1998


RESULT 15
A71622
hypothetical protein PFB0155c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71622
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71622
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-507 <GAR>
A;Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AAC71821.1; PID:g384511
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0155c

  Query Match             11.0%;  Score 85;  DB 2;  Length 507;
  Best Local Similarity   29.8%;  Pred. No. 29;
  Matches   28;  Conservative   17;  Mismatches   31;  Indels   18;  Gaps    4;

Qy          66 DQYNNSSSLRNHLNNLEDIMAQISIT------------NGDKE---VEDYEKNIKKARN 109
               |:|:| |  : ::|| : |  ::||              | |||   ::   ||: |: |
Db          53 DEYDNISRSKENINNSKKIKNELSIKDNMHDYIYDDRIYNNDKEKNVIKSDNKNVIKSDN 112

Qy         110 KLRVIASITKETWKIDSLDNLVKEVGGTLQVSEN 143
               |  || |  |   | |:  |::|     :  |:|
Db         113 K-NVIKSDNKNVIKSDN-KNVIKSDNKNVIKSDN 144


Search completed: February 16, 2004, 14:14:46
Job time : 20 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 09:07:46 ; Search time 2877 Seconds
(without alignments)
6356.136 Million cell updates/sec

Title: US-09-980-054A-11
Perfect score: 447
Sequence: 1 atgtcagatatagatataga.....accccgatgatatgtggtga 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	447	100.0	447	6	AX057345	AX057345 Sequence
	2	442.2	98.9	447	6	AX240813	AX240813 Sequence
	3	442.2	98.9	447	6	AX488762	AX488762 Sequence
	4	324.4	72.6	326	6	AX057353	AX057353 Sequence
c	5	71.4	16.0	253305	3	PFMAL3P7	AL034559 Plasmodiu
c	6	70.8	15.8	863	11	CNS06EVQ	AL395628 T7 end of
c	7	70.4	15.7	4865	3	AF364131	AF364131 Anopheles
	8	69.2	15.5	67970	3	PFMAL1P3	AL031746 Plasmodiu
c	9	69.2	15.5	178193	2	AC068949	AC068949 Homo sapi
	10	68.8	15.4	174835	2	BX248097	BX248097 Danio rer
c	11	67.6	15.1	94434	5	AL929250	AL929250 Zebrafish
c	12	67.2	15.0	170627	2	AC125567	AC125567 Rattus no
c	13	66.4	14.9	110000	2	PFMAL7P1_04	Continuation (5 of
c	14	66.4	14.9	286208	2	AC117140	AC117140 Rattus no
	15	65.8	14.7	25753	8	AF437291	AF437291 Saccharom
	16	65.8	14.7	110000	3	AC116984_3	Continuation (4 of
	17	65.6	14.7	27491	9	AC010284	AC010284 Homo sapi
	18	65.4	14.6	271546	3	AE014843	AE014843 Plasmodiu
c	19	65	14.5	65	6	AX485584	AX485584 Sequence
c	20	65	14.5	840	8	CNS018QK	AL110675 Botrytis
c	21	65	14.5	214955	2	BX511081	BX511081 Danio rer
	22	65	14.5	254449	3	AE014817	AE014817 Plasmodiu
c	23	64.8	14.5	119733	9	HSDJ655C5	AL121956 Human DNA
c	24	64.8	14.5	250029	3	AE014820	AE014820 Plasmodiu
	25	64.6	14.5	58437	9	BX276094	BX276094 Human DNA
	26	64.6	14.5	181845	2	BX005330	BX005330 Danio rer
c	27	64.6	14.5	347582	3	PFMAL4P1	AL034557 Plasmodiu
	28	64.4	14.4	153751	3	AC116551	AC116551 Dictyoste
	29	64.4	14.4	330050	3	PFA929355	AL929355 Plasmodiu
c	30	64.2	14.4	1104	8	SCYNL260C	Z71536 S.cerevisia
	31	64.2	14.4	2961	8	YSCATX1HF	L35270 Saccharomyc
c	32	64.2	14.4	33016	8	SCCHXIVL	X96722 S.cerevisia
c	33	64.2	14.4	182870	3	AC116960	AC116960 Dictyoste
	34	64.2	14.4	246611	2	AC111404	AC111404 Rattus no
c	35	64.2	14.4	340552	3	PFA929354	AL929354 Plasmodiu
	36	64.2	14.4	349980	6	AX344564	AX344564 Sequence
c	37	64	14.3	51014	2	AC011856	AC011856 Homo sapi
c	38	64	14.3	110000	3	PFMAL1P2_1	Continuation (2 of
c	39	64	14.3	152900	5	AL954652	AL954652 Zebrafish
	40	64	14.3	303671	1	AE015937	AE015937 Clostridi
c	41	63.8	14.3	110000	3	AC116984_0	AC116984 Dictyoste
c	42	63.8	14.3	156035	5	BX005364	BX005364 Zebrafish
	43	63.8	14.3	347050	3	PFA929351	AL929351 Plasmodiu
	44	63.6	14.2	110000	3	AC116957_1	Continuation (2 of
c	45	63.4	14.2	115866	2	AC135614	AC135614 Pan trogl

ALIGNMENTS

RESULT 1
AX057345
LOCUS AX057345 447 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 11 from Patent WO0075305.
ACCESSION AX057345
VERSION AX057345.1 GI:12310086
KEYWORDS .
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1
AUTHORS Lalanne,J.L. and Rocher,C.
TITLE Novel candida albicans genes and proteins coded by said genes
JOURNAL Patent: WO 0075305-A 11 14-DEC-2000;

```
          HOECHST MARION ROUSSEL (FR)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
     CDS             1. .447
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=12
                     /protein_id="CAC22438.1"
                     /db_xref="GI:12310087"
                     /translation="MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTG
                     FQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEK
                     NIKKARNKLRVIASITKETWKIDSLDNLVKEVGGTLQVSENPDDMW"
BASE COUNT      192 a     38 c     82 g    135 t
ORIGIN

  Query Match           100.0%;  Score 447;  DB 6;  Length 447;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-55;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447


RESULT 2
AX240813
LOCUS       AX240813                447 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 51 from Patent WO0160975.
ACCESSION   AX240813
VERSION     AX240813.1  GI:15797749
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 0160975-A 51 23-AUG-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      193 a     38 c     80 g    136 t
ORIGIN

  Query Match           98.9%;  Score 442.2;  DB 6;  Length 447;
  Best Local Similarity  99.3%;  Pred. No. 5.6e-55;
  Matches  444;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447


RESULT 3
AX488762
LOCUS       AX488762                447 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6062 from Patent WO02053728.
ACCESSION   AX488762
VERSION     AX488762.1  GI:22322774
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6062 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      193 a     38 c     80 g    136 t
ORIGIN

  Query Match           98.9%;  Score 442.2;  DB 6;  Length 447;
  Best Local Similarity  99.3%;  Pred. No. 5.6e-55;
  Matches  444;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447


RESULT 4
AX057353
LOCUS       AX057353                 326 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 19 from Patent WO0075305.
ACCESSION   AX057353
VERSION     AX057353.1  GI:12310094
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Lalanne,J.L. and Rocher,C.
  TITLE     Novel candida albicans genes and proteins coded by said genes
  JOURNAL   Patent: WO 0075305-A 19 14-DEC-2000;
            HOECHST MARION ROUSSEL (FR)
FEATURES             Location/Qualifiers
     source          1. .326
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
BASE COUNT      146 a      27 c      53 g     100 t
ORIGIN

  Query Match             72.6%;  Score 324.4;  DB 6;  Length 326;
  Best Local Similarity   99.7%;  Pred. No. 6.6e-38;
  Matches  325;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         12 AGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGG 71
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 AGATATAGATAATGTATTAAATTTAGAAGAAGATCAATATGAATTAGGATTTAAAGAAGG 60

Qy         72 TCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGG 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGG 120

Qy        132 ATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCA 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCA 180

Qy        192 TATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATAT 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATAT 240

Qy        252 TATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATAT 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATAT 300

Qy        312 TAAAAAGGCAAGAAATAAATTAAGAG 337
              |||||||||||||||||||||||||
Db        301 TAAAAAGGCAAGAAATAAATTAAGAG 326


RESULT 5
PFMAL3P7/c
LOCUS       PFMAL3P7                 253305 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum MAL3P7, complete sequence.
ACCESSION   AL034559 AL008974 AL008975 AL008981 AL008983 AL009015 AL010138
            AL010143 AL010146 AL010154 AL010157 AL010160 AL010165 AL010169
            AL010187 AL010189 AL010207 AL010208 AL010214 AL021885 AL021888
            AL139179 AL844502 Z98556 Z98557 Z98558
VERSION     AL034559.4  GI:8052273
KEYWORDS    HTG; 40S Ribosomal protein S3A; acyl transferase;
            acylaminoacyl-peptidase; ATP-dependent RNA Helicase; cyclophilin;
            elongation factor; F49C12.11-like protein; HesB-like domain
            protein; histone H2A variant; kinesin-related protein;
            N-acetylglucosamine-1-phosphate transferase; protein kinase; R-CG7
            repeat; R-FA3 repeat; rep11; rep20; rifin; RNA-binding protein;
            stevor; T-complex protein 1 epsilon subunit; telomere; var.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253305)
  AUTHORS   Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
            Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,
            Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
            Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
            Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
            Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
            Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
            Barrell,B.G.
  TITLE     The complete nucleotide sequence of chromosome 3 of Plasmodium
            falciparum
  JOURNAL   Nature 400 (6744), 532-538 (1999)
  MEDLINE   99376085
   PUBMED   10448855
REFERENCE   2
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   3  (bases 1 to 253305)
  AUTHORS   Lawson,D., Bowman,S. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On or before May 14, 2001 this sequence version replaced
            gi:2982540, gi:2982541, gi:2982544, gi:2894453, gi:2894464,
```

```
        gi:2982551, gi:2982556, gi:2982558, gi:2982563, gi:2982564,
        gi:2982566, gi:2894491, gi:2982569, gi:2695974, gi:2894502,
        gi:2673766, gi:2894496, gi:2982577, gi:2894588, gi:2894589,
        gi:2982538, gi:2982539, gi:2894380, gi:4725992.
        For more information about this sequence or the Malaria Project,
        see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .253305
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="3"
                     /clone="MAL3P7"
     gene            458. .4060
                     /gene="MAL3P7.1"
                     /note="synonym: PFC0860w"
     CDS             458. .4060
                     /gene="MAL3P7.1"
                     /note="Similarity: to kinesin-related protein.
                     D.melanogaster kinesin-like protein 67A (TR:P91945) BLAST
                     Score: 664, sum P(2) = 1.9e-67; 28% identity in 707 aa
                     overlap; Pfam: match to PF00225 kinesin (6.1e-86), Kinesin
                     motor domain Score 311.54 E-value 6.1e-86"
                     /codon_start=1
                     /product="kinesin, putative"
                     /protein_id="CAB39023.1"
                     /db_xref="GI:4493964"
                     /db_xref="GOA:O97277"
                     /db_xref="SPTREMBL:O97277"
                     /translation="MRKTKNLNDPFFLNNYYYNKEENNPIIICSDEKERKTKIYSTNH
                     HTNNILNYNDHTLKLICNITENNKAILHSNKSNLQKNNILMPSYMQKKGTHIRETIKN
                     VYPNINGEPSTSVENITNGEHFINGQYDALKNMSLNNYDHQHNNIMNNISNKNKLFVL
                     NNNNNNNNQDVIENMNEYPITSKNIYDSIYIPQINIKNIINSEEINNNNNNINDNNNH
                     NYTNIHNNNNNMNSNPTSSVTSKKNENNNLINTLNAYSNVKVAVRIKPIGESEENIVS
                     IFNKNYVLIEKENEKECYLLSQKKKQSTYVFDSVFDVNATQEEVFFQTAKPLIPHVFK
                     GINCTVFAYGATGSGKTYTMLDDKNQNGIVQLSLLELFTIINEKKCRNIKVLMSFLEV
                     YNETIRDLLGKEKNKTLEVQEDVAEVKVSNLCEIEVNNYEQAMLLINEGVKNRKMSPT
                     RANKVSSRSHAILQIYVYNEILDDNMNTISYKAKLCFVDLAGSERASATSNKGERFKE
                     GSYINQSLLALANCINSLASNRNISKVRVKYRDSKLTHLLKNSLEGNCLVVMIANINP
                     SRTSFQESNNTLKYAFRARNIKLCATVQTNDNKESDIEKILKKNENLQKEYDTLLGKY
                     TNLKEFFFIINVINQLYKKQISCYKLIENISDNMSSMELKQDITMYDQLVKMKSDEYR
                     KKVDSLKDLYQEEKQFLNNLFDTFLEKNLNYVINSKDVNDNNKSLLEEMIFFKHNENK
                     VNENFLVNEKVVDKNNVLNGNVMVDENVVDKNNVLNGNVMVDENVMVDENVMVDENVM
                     VDENVMVDENVMVDENVMVDENVMVDENVMVDENVMVNENVVVDKNVSVNENIVVSEKHKVGLS
                     AEGKSESHNKNNKDDIEDNDKDTIKDIHNNNNSSDNNDDEYQSANSPVESDIVKKEKK
                     KKIPINMETKKKRTMNGTKDPIHKTPYDINIVGILNKEDVSNKSNDYNTNKNIEKNNY
                     EKKGEYNPFHNNLTDMQNSILYNIINNNVENSPHSPRMKKNVAKMLLKGNLNTANFIL
                     DDDTIKNMNSNKISDKHNMKSNNILNNENGKINDKSKKCKNINNNNNNNNNNNNNNNN
                     NNNNNNNSSSSSSGKVDGINILNNSNTNERLHTFSGVYSLNLNDEIKIEINKKDMEKN
                     DIHLTSIDTISKIHARDLLKENKRKLENFQENIKHEHKDEVSLYVKKKKIKKKM"
     gene            join(5457. .5577,5689. .5787,5923. .6077,6189. .6743)
                     /gene="MAL3P7.2"
                     /note="synonym: PFC0865w"
     CDS             join(5457. .5577,5689. .5787,5923. .6077,6189. .6743)
                     /gene="MAL3P7.2"
                     /note="Similarity eg to C.elegans RNA-binding protein
                     (TR:Q18318) BLAST Score: 378, sum P(1) = 3.8e-35; 37%
                     identity in 297 aa overlap; splicing conserved in P.
                     knowlesi and P. yoelii; Pfam: match to PF00076 rrm, RNA
                     recognition motif. (a.k.a. RRM, RBD, or RNP domain) Score
                     58.01, E-value 1.1e-20"
                     /codon_start=1
                     /product="RNA-binding protein, putative"
                     /protein_id="CAB39067.1"
                     /db_xref="GI:4494008"
                     /db_xref="GOA:O97318"
                     /db_xref="SPTREMBL:O97318"
                     /translation="MGLFDKIRNIEKLNEAELKNIGNNDSSWHDQYRDSSYIYIGNLD
                     NRLTEGDIVIVFSQYGEPIDVNLVRDNETGKSKGYCFLSYADQRSTILAVDNFNGYKL
                     LERPLVVDHILNYRLPKKYLKDADKNEYKPTGAEGQGIGVYNVVESEIKLSKVFDKIK
                     NKSNEEKKKKLLDEDELWALNFEKSIKKDIISPIGHDEKSRHNEGMKEEEEDEDDEDD
                     DDEDDDDDDSVDIKYKRHKEKRKSLTTKKYDKKEKHKRKSDHRDKHRRRENHSRHREK
                     EKDKKSHKRRHKHSYDKYSSRSRSYSTSSSTDR"
     gene            join(7801. .7928,8084. .8234,8392. .8583)
                     /gene="MAL3P7.3"
                     /note="synonym: PFC0870w"
     CDS             join(7801. .7928,8084. .8234,8392. .8583)
                     /gene="MAL3P7.3"
                     /note="Similarity to P. falciparum elongation factor
                     1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E():
                     1.1e-15, 53.2% identity in 141 aa overlap, revised:
                     shortened exon 2, splicing confirmed in P. knowlesi; Pfam:
                     match to PF00736 EF1BD, EF-1 guanine nucleotide exchange
                     domain Score 61.10, E-value 7.9e-22"
                     /codon_start=1
                     /product="elongation factor 1 (EF-1); putative"
                     /protein_id="CAB39068.2"
                     /db_xref="GI:8052274"
                     /db_xref="GOA:O97319"
                     /db_xref="SPTREMBL:O97319"
                     /translation="MANTYDELYVPLSYYILQNEGGNTSKIDQANTKKPKKKEVINKS
                     SLIIDIKPYGENTDLDEVLKLVKNITMEGLTWGKAHKKTPFAFGLFKLQVSCVIVDDL
                     VNTDELIETIENLGLDNEQLQKKKQMDDDEENYDEDDEIGGLVQSAEIISFNKL"
     gene            join(10596. .10737,11044. .11146,11373. .20416,20578. .20690)
                     /gene="MAL3P7.4"
                     /note="synonym: PFC0875w"
     CDS             join(10596. .10737,11044. .11146,11373. .20416,20578. .20690)
                     /gene="MAL3P7.4"
                     /note="Revised: added 3' exon, possibly spliced at 5' end,
                     revised: added 2x 5' exons by similarity with P. knowlesi;
                     signal anchor predicted by SignalP 2.0 HMM (Signal peptide
                     probabilty 0.001, signal anchor probability 0.997) with
                     cleavage site probability 0.000 between residues 36 and 37
                     Pfam match to entry PF00005 ABC_tran, ABC transporter,
                     score 72.50, E-value 9e-18"
                     /codon_start=1
                     /product="transporter, putative"
                     /protein_id="CAB39024.3"
                     /db_xref="GI:23477031"
                     /db_xref="GOA:O97278"
                     /db_xref="SPTREMBL:O97278"
                     /translation="MKIRKLYGLLKKIYYEKRKDFFLYFFFLLIPFLLISFHLFLRNI
                     SDKYSLEIYNTSSPNDKINLNDTIRQYVLFLSNELCSKEINDEIYVNHICLTPGNELM
                     KDFLSYAENNDLNIFKLYNSEEDCLKNLKYAIELKSEDFKNVNSLELNDDILNIHEND
                     KNSLLYNLHTNIINLKSSEFNINDLYKDVYTENEFENIKNLKINPHILKVVKNTEFYK
                     NFIKDIKNLKNENHYTQYFNDDRKKLFFYNFVKNNLVETKYSCGLLSVENVNYYKKQK
                     NVKFSYLFGLSPQSEGFKGKGLYKNAPLYNKYENKQYGNPFHLYYPYVTLESDKKINE
                     KNERNEKNERNEKNERNEKNEKNGKNEKNKDILSSTNNNVVMKNGIQSYHNNNSINSN
                     DDQKKKNNNNNNYYYYNNPDGLITNVKYKIRVGDYALLNSNEKYFFNDININLKQNFI
                     NFNTVDDLSFNIYFNEWYYFSFFIVLEYQFNSFILNYNADILKKNNLLTPRYDEINEL
                     NKEREITMEYKKTDEKEYQVNNEYINENVGYDEKGGKYKTELKLNDFFIYKMPMKSMK
                     INAFDTFEKNIFRVVVFLCVCLFIVNICFDINKERKINIENFLCCLKINKYYYYYSWL
                     LFYFIILFFYNIIFTLVVYFYVYKCLMNYFILFFLIYFFLMNSLLFTIICMQFSNNSS
                     INYIATFLLFFLFSSFRLIIHSGASNILTFFVLLIPHSSFCLSLDFIFILIKNNIKID
                     YKQLFIKFENISLMHLIICSIVSFVLLICILNYIIYYKRKKMNMLNTKYRGKSKSITN
                     DKRKKSQQKTYEYGNINYDMSEKYPTHQLKDNSNDFMLEGSDEDDTCYELDTLHKKQV
                     NSQMCLRIEKIRNDRRESNGHPYDIPMIPLRHNNNINNNINNNINNNINNNINNNINN
                     NQHRNNQCYLIIQNVNKFYGKKHALKDVSLTLRSNRIFVLLGENGSGKSTLINIITKM
                     ISKDSGEINFFKNSYKNKKKDIYMDVNKNRRNGLCGRICSNGYIKEFVMNRIFKRNCN
                     DTNNYDNNSNNNNNNDNHNDNHSNNSGINSSFNNNNNVHNKNIKNDLEISYCSQNVIL
                     YENLTFYETIKIFLLYYNKNVDKYLNKKRTLKIMNDLDLVQYLNDQIKNLMDEVKKKI
                     SIFICFLVKRDIYILDEPFIALDIKTKTKLFKFFDKIKKNNIIFICTHDIYEANNFAN
                     DIAVIKSGQIIFKGSKNSFQKLIEYKFTLNIQFNGLPNEEQTNYLSQEKIMSALNKKG
                     VIKSPRKVANNYVDGYNVGRISSHESSINDKKKDNNDNNNDDDNNKNNSVDHIDHLFD
                     ILVTNINEFTNNRDIIKKNIIKFIKGTRNENKNCFIFFNENHIYCTYKINELESLKKL
                     LCVLNKFKNILTYQLKTIDIYYTYIYIYTLNEKKKLLKNIQDKDIKYLIEIDPLFFLF
                     FQNFKYFNELNNLLLMKNNHMQPISYNFAYLRDILVRGNENKNIINSTNHDDNPYDIN
                     MYNNHNKYGKYKNNNNNNSFYNMRDDTELTDIEENISSKKKKNKFIEKEFSYNTYRNN
                     KNNDENDEYGEYDNYDDNMIMSNTSNIMKKKNIKKQNIFKTCINFFTTYIKPTLLLKL
```

```
Query Match          16.0%;  Score 71.4;  DB 3;  Length 253305;
Best Local Similarity  51.5%;  Pred. No. 0.045;
Matches  191;  Conservative    0;  Mismatches  176;  Indels    4;  Gaps    1;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
```

```
Db      222778 AGAAAAAAATGAAAAAAAAGAAAAAAAGAAAAATGAAATATAAAAAAAAAATTTATTAAAA 222719

Qy          66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125

Db      222718 TAAAAAAAAAAAAAAAAAAAAAAAAGAAAAATATTTTAAAAATAATAAAAATTATAATAA 222659

Qy         126 AACTGGA----TTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTT 181

Db      222658 AAATATAAATTTTGATAAAATAAAAAATGAAAAATATTATCAAAAAAAAATTAAAAAAAA 222599

Qy         182 GGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATT 241

Db      222598 ATTTTATATAAAAAAACAATGATTATAAAAAAAATTTATTACATAAAAAAAAAAGAAAA 222539

Qy         242 TGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATG 301

Db      222538 AAAAACATGTTAAAAAAAAAATATATATCATAAAATAAAAAAAAAAAGTTAAAAAATATT 222479

Qy         302 AAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAA 361

Db      222478 AAAATAAAAATATATATCATAAAATAAAAAAAAATTAAAAAATTTATTTAAATAAAAAAA 222419

Qy         362 CTTGGAAAATT 372

Db      222418 AATTAAAAAAT 222408
```

RESULT 6
CNS06EVQ/c

LOCUS	CNS06EVQ 863 bp DNA linear STS 10-JAN-2001
DEFINITION	T7 end of clone AR0AA025B09 of library AR0AA from strain CBS 732 of Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION	AL395628
VERSION	AL395628.1 GI:12147375
KEYWORDS	STS.
SOURCE	Zygosaccharomyces rouxii
ORGANISM	Zygosaccharomyces rouxii Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE	1 (bases 1 to 863)
AUTHORS	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE	20584711
PUBMED	11152876
REFERENCE	2 (bases 1 to 863)
AUTHORS	de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B., Wincker,P., Artiguenave,F. and Souciet,J.
TITLE	Genomic exploration of the hemiascomycetous yeasts: 8. Zygosaccharomyces rouxii
JOURNAL	FEBS Lett. 487 (1), 52-55 (2000)
MEDLINE	20584718
PUBMED	11152883
REFERENCE	3 (bases 1 to 863)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES	Location/Qualifiers
source	1. .863 /organism="Zygosaccharomyces rouxii" /mol_type="genomic DNA" /strain="CBS 732" /db_xref="taxon:4956" /clone="AR0AA025B09" /clone_lib="AR0AA" /note="end : T7"
misc_feature	<3. .>155 /note="similar to Saccharomyces cerevisiae ORF YGR184c [UBR1 ; ubiquitin-protein ligase]" /evidence=not_experimental
BASE COUNT	196 a 105 c 60 g 417 t 85 others
ORIGIN	

```
  Query Match           15.8%;  Score 70.8;  DB 11;  Length 863;
  Best Local Similarity   42.1%;  Pred. No. 0.24;
  Matches  166;  Conservative   45;  Mismatches  177;  Indels    6;  Gaps    1;

Qy          35 TAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTCAAATACAAGGAACAAAAGATC 94

Db         714 TAAATGAATAAAATAATTAAATAATAATGAAAAAAAAAAAAAGAATAAATAAAAAAAWATW 655

Qy          95 AATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTTCAACGATTTTTAATCATTG 154

Db         654 ATWATTAATATATATAWAAATATATATAWAAAAAATAATATAWATAAATAWTTATAWGAG 595

Qy         155 GTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATAGATCAATATAATAACTCTT 214

Db         594 AAGWAAWAGTAWWAAATWTAAAGAAAACAATAAAAAAAWTAAWWAAATWWGAGTGGGTTA 535

Qy         215 CTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATGGCACAAATTTCTATAACGA 274

Db         534 AWAAAATTAWAAAWAWGAWAAAAAATAAAGAAAATAGTA------ATATTAATCWRAAWA 481

Qy         275 ATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAAAAGGCAAGAAATAAATTAA 334

Db         480 ARAAAAWAAATTAAGTTGTGAGTWGARAGAAAAAAAWTGARRTGGRAAAAAAWRTAATAA 421

Qy         335 GAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTGATTCATTGGATAATTTGGTGA 394

Db         420 AGGWAAGGAAAAATRARAAWAATATAAAWWRAWAAAAWWAAAGAAGAAGWTTTGTARAWT 361

Qy         395 AAGAAGTAGGTGGAACTTTACAAGTTAGTGAAAA 428

Db         360 AWAAAAAAATTRAARGAAAMAAARAAAGTTAAAA 327
```

RESULT 7
AF364131/c

LOCUS	AF364131 4865 bp DNA linear INV 27-NOV-2001
DEFINITION	Anopheles gambiae putative odorant receptor Or2 gene, complete cds.
ACCESSION	AF364131
VERSION	AF364131.1 GI:17105032
KEYWORDS	.
SOURCE	Anopheles gambiae (African malaria mosquito)
ORGANISM	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE	1 (bases 1 to 4865)
AUTHORS	Fox,A.N., Pitts,R.J., Robertson,H.M., Carlson,J.R. and Zwiebel,L.J.
TITLE	Candidate odorant receptors from the malaria vector mosquito Anopheles gambiae and evidence of down-regulation in response to blood feeding
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (25), 14693-14697 (2001)
PUBMED	11724964
REFERENCE	2 (bases 1 to 4865)
AUTHORS	Fox,A.N., Pitts,R.J., Robertson,H., Carlson,J. and Zwiebel,L.J.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) Biological Sciences, Vanderbilt University,
            390 24th Ave. S., Nashville, TN 37235, USA
FEATURES             Location/Qualifiers
     source          1. .4865
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7165"
     mRNA            join(<455. .670,768. .851,948. .1372,1571. .1670,1753. .1854,
                     1963. .2070,3429. .>3530)
                     /product="putative odorant receptor Or2"
     CDS             join(455. .670,768. .851,948. .1372,1571. .1670,1753. .1854,
                     1963. .2070,3429. .3530)
                     /note="G protein-coupled receptor"
                     /codon_start=1
                     /product="putative odorant receptor Or2"
                     /protein_id="AAL35507.1"
                     /db_xref="GI:17105033"
                     /translation="MLIEECPIIGVNVRVWLFWSYLRRPRLSRFLVGCIPVAVLNVFQ
                     FLKLYSSWGDMSELIINGYFTVLYFNLVLRTSFLVINRRKFETFFEGVAAEYALLEKN
                     DDIRPVLERYTRRGRMLSISNLWLGAFISACFVTYPLFVPGRGLPYGVTIPGVDVLAT
                     PTYQVVFVLQVVLTFPACCMYIPFTSFYATCTLFALVQIAALKQRLGRLGRHSGTMAS
                     TGHSAGTLFAELKECLKYHKQIIQYVHDLNSLVTHLCLLEFLSFGMMLCALLFLLSIS
                     NQLAQMIMIGSYIFMILSQMFAFYWHANEVLEQSLGIGDAIYNGAWPDFEEPIRKRLI
                     LIIARAQRPMVIKVGNVYPMTLEMFQKLLNVSYSYFTLLRRVYN"
BASE COUNT     1297 a    1024 c     954 g    1590 t
ORIGIN

  Query Match           15.7%;  Score 70.4;  DB 3;  Length 4865;
  Best Local Similarity  51.0%;  Pred. No. 0.18;
  Matches  196;  Conservative    0;  Mismatches  181;  Indels    7;  Gaps    1;

Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              | | ||  | |   | || |   || | | | | || || || |||| |||| |  ||
Db       2819 TAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAAGAATAATAATAA 2760

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
              || ||    ||  | ||  |  | || |  |||| | | ||  |||  | |||  | ||
Db       2759 TAATAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTATTATTATGAT 2700

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
              | | || ||   || ||| |  |||        ||| | |  ||| |    || ||
Db       2699 AATAATGTATAATAATGAATAACAATA-------ATAATAATTAATAATAATAAATTAAT 2647

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              | ||     ||| || |  ||||||| | || ||    |   ||| ||   ||||||
Db       2646 TAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAATAATAATAATAATAAT 2587

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              |  ||| ||      || | || ||  | |||    ||||  |   | |  || || | |
Db       2586 AATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATAATAATAAT 2527

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              || |||| ||| ||    |  ||||||   |          || ||  | |||  | | |
Db       2526 AATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATAATAAT 2467

Qy        364 TGGAAAATTGATTCATTGGATAAT 387
                || | ||    |  |||||
Db       2466 AATAATAATAATAATATTAATAAT 2443


RESULT 8
PFMAL1P3
LOCUS       PFMAL1P3                67970 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum DNA from MAL1P3.
ACCESSION   AL031746 AL844501
VERSION     AL031746.9  GI:6594243
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   2
  AUTHORS   Murphy,L., Pain,A., Berriman,M., Bowman,S., Churcher,C., Harris,B.,
            Harris,D., Lawson,D., Quail,M., Rajandream,M., Hall,N. and
            Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Dec 16, 1999 this sequence version replaced gi:5763807.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .67970
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="1"
     gene            complement(join(201. .257,455. .586,748. .804))
                     /gene="PFA0575c"
                     /note="synonym: MAL1P3.01b"
     CDS             complement(join(201. .257,455. .586,748. .804))
                     /gene="PFA0575c"
                     /note="Similar to Schizosaccharomyces pombe hypothetical
                     protein c119.18 spbc119.18 SWALL:Q96VG1 (EMBL:AL022117)
                     (69 aa) fasta scores: E(): 1.3e-06, 39.68% id in 63 aa,
                     and to Homo sapiens protein 15e1.1 15e1.1 SWALL:15E1_HUMAN
                     (SWALL:O43715) (76 aa) fasta scores: E(): 0.00046, 38% id
                     in 50 aa, and to Mus musculus protein 15e1.1 15e1.1
                     SWALL:Q9D8Z2 (EMBL:AK007514) (76 aa) fasta scores: E():
                     0.00054, 38% id in 50 aa"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="CAD49075.1"
                     /db_xref="GI:23477009"
                     /translation="MENRSKTISQNTIKAHVEANDECKEKKEKYLKCFNNWYKNNFLK
                     GDLTQACDDYYEDYQICVLNDLNKKGLGHLSNVEKEK"
     gene            complement(join(1392. .1461,1752. .2598,2748. .2848,
                     2990. .3276))
                     /gene="PFA0580c"
                     /note="synonym: MAL1P3.01"
     CDS             complement(join(1392. .1461,1752. .2598,2748. .2848,
                     2990. .3276))
                     /gene="PFA0580c"
                     /note="Originally annotated as Plasmodium falciparum;
                     conserved hypothetical protein but pfam match and other
                     evidences suggest that it is a putative deoxyribonuclease;
                     earlier start site possible to make a 412 aa protein
                     Pfam match to entry PF01026 TatD_DNase, TatD related
                     DNase, score 117.30, E-value 2.9e-31
                     Similar to Plasmodium falciparum conserved hypothetical
                     protein, upf0006 family mal1p3.01 SWALL:Q9U0N6
                     (EMBL:AL031746) (412 aa) fasta scores: E(): 3.7e-150, 100%
                     id in 412 aa, and to Saccharomyces cerevisiae putative
```

```
                deoxyribonuclease ybl055c ybl055c or ybl0512 or ybl0511
                SWALL:YBF5_YEAST (SWALL:P34220) (418 aa) fasta scores:
                E(): 2.5e-19, 30.47% id in 397 aa"
                /codon_start=1
                /product="TatD-like deoxyribonuclease, putative"
                /protein_id="CAD49076.1"
                /db_xref="GI:23477010"
                /translation="MKLVFHYIKYINVLFYISIIFLKSNSLKIYNDLRYISTVNKYKV
                LQIKKRSNLKKNHNIRKMEDNESSFIDIGSNLTDKMFDGVYNSKKHENDLQNVLNRAK
                NNNVDKIIITCTCLAEIDKSLKICETYDPEGKFLYLSAGVHPTNCYEFIDKNKHEEKE
                IIAKKEYEEFIKYFKNEQVENSKMENGNKKICDGEKDMNNLNEILLEKNLDTIPGFKY
                NEKDKEYLENLKNKIIKYPNRIVCIGEIGLDFDRLYFCSKYIQIKYFIFQLKLVQMFN
                LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVIHSFTDKEDIVHIIVQNYKNLYIGVNG
                CSLKSLENINAVKKIPLNLLLLETDAPWCGVKKTHASYEYIKDTYEKRAYTNLKKIKN
                IIKCDDNTIFKERNEPYNIADIAEITYKVREEAVPFDLFCKK"
gene            5005. .5496
                /gene="MAL1P3.02"
                /note="synonym: PFA0585w"
CDS             5005. .5496
                /gene="MAL1P3.02"
                /note="contains possible signal sequence. Signal peptide
                predicted SignalP 2.0 HMM (Signal peptide probabilty
                0.640, signal anchor probability 0.357) with cleavage site
                probability 0.504 between residues 27 and 28. ScanRegExp
                hit to PS00867, Carbamoyl-phosphate synthase subdomain
                signature 2"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB63557.1"
                /db_xref="GI:6594245"
                /db_xref="SPTREMBL:Q9U0N5"
                /translation="MKLLNNRFVVLCPIIILFFFLNSVVLGNNNRNNINFHETENAAK
                AMRKLLSGEINSIKLDNGDELKIKLNDEKHKDSTKWDKSYSFISNLEEEKYSQTDLFR
                KKQEINEANTKIIEDRQEFYILNNDEIENIATRFVLENNFDELYIQSFKQSLIDIIQS
                LNN"
misc_feature    8179. .10303
                /note="Putative centromere"
gene            14884. .20352
                /gene="MAL1P3.03"
                /note="synonym: PFA0590w"
CDS             14884. .20352
                /gene="MAL1P3.03"
                /note="Putative ABC transporter
                Pfam match to entry PF00005 ABC_tran, ABC transporter,
                score 45.00, E-value 1.7e-09; HMMSmart hit to SM00382,
                ATPases associated with a variety of cellular activities;
                ScanRegExp hit to PS00211, ABC transporters family
                signature
                Similar to Schizosaccharomyces pombe ATP-binding cassette
                transporter abc1 abc1 or spac9e9.12C SWALL:ABC1_SCHPO
                (SWALL:Q92337) (1427 aa) fasta scores: E(): 4.2e-08,
                21.46% id in 1146 aa"
                /codon_start=1
                /product="ABC transporter', putative"
                /protein_id="CAB63558.1"
                /db_xref="GI:6594246"
                /db_xref="GOA:Q9U0N4"
                /db_xref="SPTREMBL:Q9U0N4"
                /translation="MTTYKENVGISNKGNKKKKSCQNISFLNFLSFDWIRPLINDLIK
                GDIQELPNICRNFDVPYYASKLEENLRDIEVEDSEFYSEKNSSNEHVLHHCNSNDASE
                KKVYNVYYHNILWSILKTFKFRIILIISFYILETLIVTLGGKFIDYYMRILEGQKIPV
                YISFLKDFKVFSGLVVVMIMFFHLFFEALLHFYFHLFTINLKVSLMYFLYKINLCSNN
                NHLQNPDAFYNTYRKFSSQTEIDEISRDFLSIGKNASSSSSGIKNNNKNIDNNKFVEN
                DYIINFIKSTKKMEKDSLNENRSLPNVNIYNIMFSDVPSVTFFVTSCINLFNVFVKIF
                MSFYVFHIKIGSNSVGIAIWLSIALYSAMILFEFLPSLFKSKYLIYRDKRIDNMHHVL
                KEFKLIKMFNWESFAFKYINIFRMKEMKYCKIRLYLSNIGVFISSISSDIVEVVIFFI
                YLKDRLNKKEEIKFTSIIMPLYVYKILISNVANFPNLVNNVMEGIVNIKRLNNYINDH
                LYYNDIKNYFMYRTRYNEDYNIVVDKTFLQNENITSHDDGTSHNLKHLKNVIKNKLTN
                MFKYFFFYHKMNYHKNIINKQILSGLLKNVDDNTNKKICFQEHKSNSTYNYNSSHIHE
                KKEEYENIHNSSNSTMSNEFKEKKKNNEYIIKLENCSFGLSYDNKCDNDHILKNINFN
                LKRNSLAIIIGNVGSGKSAFFHSILGDFNMTHGNLYIENFFKKMPILYVPQNSWLFMG
                NIRSMILFGNEYNPLIYKYTILQSELLNDLSTIEHGDMKYINDDHNLSKGQKVRICLA
                RALYEHYIHMHKLCTDYEKKLIQPNEILDKDLINNKNISSYNNKKSKLVNYNIPFNEN
                YLQKCLMDDNNFYLYLLDDIFTSLDPSISKKIFSNLFCKEDNISFKDNCSFIISMNKS
                TLDNFLIEDILDNVQYEVNIFEIQDKTLKYRGNISEYMEKNNLNITKESHWGYSNLNT
                IDYTRIKLFDEVELNHVKHSNKMIYKEAYFVKGNTESVSFEIDSINKEYIKKMKKKNY
                KKEHMNKNNKDNNNNNNNSNKDDHININMNDNHRNYNDINLGPNSTDDSPTVSSLGNE
                YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMPFVKSSSNMINNPSNFKYEDNSSS
                FKGSISLETYLWYFQQVGFVLLTSVVIFMLISIFTDEIKFVFLTMMSIISKNNKEHSD
                TILQKQVRYLEYFVILPIISLVTSGICFSMIIYGNITSAIKVHNNILYSILNAPLYIF
                YNNNLGNIINRFIIDISAFDYGFLKRIYKAFFIFFRCILSSLLIIYMIRDCIFIFPFV
                IILIYFFVFKRFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIYKKNTYHLDVYE
                HYINNFRISYFFKWLINIWASLYIKIFILLLTTYIIMHPHLYASGIIKLYKEKNYVRI
                LSTLGYCISFSARLGVIIKFLLCDYTHIEKEMCCVQRLEEFAKISNKENASMNKENEL
                NVITTQTYKEKNENISDKISAIVEYKNVSLSSIINSSQDDESKKKYGIKFENVYVSYK
                KKIPLVNGTYKYIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQG
                KITVEGRDIRTYNRKGEDSIIGILAQSSFVFYNWNIRTFIDPYNNFTDDEIVHALKLN
                GINLGKNDLYKYMHKQDMKSNYKKIIQTSKVINQSNDNTILLTNDCIRYLSLVRLYLN
                RHKYKIILIDEIPIFNLNNSVHDELNSFLIGKAKSFNYIIRNHFPNNTVLIISHHANT
                LSCCDYIYVLRKGEITYRCSYEDVKTQSELSHLLEMDD"
rRNA            23896. .26044
                /product="18s ribosomal RNA A-type"
                /note="18s rRNA exppressed in asexual stage parasites"
misc_RNA        26045. .26438
                /product="ITS1 A-type"
                /note="ribosomal operon internal transcribed spacer 1
                expressed in asexual stage parasites"
gene            26439. .26559
                /gene="mal1_5.8srNA"
rRNA            26439. .26559
                /gene="mal1_5.8srNA"
                /product="5.8S ribosomal RNA"
                /note="5.8S ribosomal exppressed in asexual stage
                parasites
                hit to vivax A-type 5.8S ribosomal percent id: 88.00 none"
misc_RNA        26560. .27435
                /product="ITS2 A-type"
                /note="ribosomal operon internal transcribed spacer 2
                expressed in asexual stage parasites"
rRNA            27436. .31539
                /product="28s ribosomal RNA (A-type)"
                /note="28s rRNA expressed during asexual development"
gene            complement(join(31966. .32476,32675. .32775))
                /gene="MAL1P3.04"
                /note="synonym: PFA0610c"
CDS             complement(join(31966. .32476,32675. .32775))
                /gene="MAL1P3.04"

  Query Match             15.5%;  Score 69.2;  DB 3;  Length 67970;
  Best Local Similarity   50.5%;  Pred. No. 0.13;
  Matches  222;  Conservative    0;  Mismatches  213;  Indels    5;  Gaps    2;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65

Db       8476 ATATAATTAAATAAATAAACTATTATATTAAATAAATTAATAATATAATTAATAAATAAA 8535

Qy         66 AGA--AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123

Db       8536 ATATTATATTAAATAAATTAAATAAAATATTAAAATATATAAATTAATAATATATATAAT 8595

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183

Db       8596 TAAATAAATAAAATATTATATTAAATAAATTAATAATATAATTAATAAATAAAATATTAT 8655

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTG---AATAAT 240

Db       8656 ATTAAATAAATTAAATAAAATATTAAAATATATAAATTAATAATATAATTAATAAATAAA 8715

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Db       8716 ATATTATATATATTAAATAAATTAAATAAAATGTTAAAATATATAAATTAATAATATAAT 8775

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Db       8776 TAAAATATTATATTAAATGAATTAAATGAAATATTATATTAAATAATTAAAATATTATTA 8835
```

```
Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              ||  | || |  |   ||| || |  |   ||| | ||  |  || | || || ||
Db       8836 TATTAAATAAATTAAATATTATATTAAATAAATAAAATAATATATTAAAATATATAAATT 8895

Qy        421 AGTGAAAACCCCGATGATAT 440
              | | || |     || ||||
Db       8896 AATAAATAAAATAATAATAT 8915


RESULT 9
AC068949/c
LOCUS       AC068949               178193 bp    DNA     linear   HTG 04-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-40N17, WORKING DRAFT SEQUENCE,
            2 unordered pieces.
ACCESSION   AC068949
VERSION     AC068949.13  GI:13540721
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178193)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178193)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Apr 4, 2001 this sequence version replaced gi:12313767.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 897
                 Center clone name: RP11-40N17
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 98% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 1% of reads
                 Chemistry: Dye-primer; 12% of reads
                 Chemistry: Dye-terminator Big Dye; 87% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 177344 bases at least Q40
                 Consensus quality: 177674 bases at least Q30
                 Consensus quality: 177760 bases at least Q20
                 Insert size: 164565; agarose-fp
                 Insert size: 178093; sum-of-contigs
                 Quality coverage: 11.2x in Q20 bases; agarose-fp
                 Quality coverage: 10.4x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9823: contig of 9823 bp in length
            *     9824     9923: gap of unknown length
            *     9924   178193: contig of 168270 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178193
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-40N17"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .9823
                     /note="assembly_name:Contig12
                     clone_end:T7"
     misc_feature    9924. .178193
                     /note="assembly_name:Contig13
                     clone_end:SP6"
BASE COUNT    56954 a  30999 c  30594 g  59483 t    163 others
ORIGIN

  Query Match             15.5%;  Score 69.2;  DB 2;  Length 178193;
  Best Local Similarity   48.9%;  Pred. No. 0.1;
  Matches  185;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
              | ||| | ||| |   | ||| || | || |  || |   ||  || ||  ||||| |
Db        572 AAATATAAATATAATTTATATAAATTATATATCAATATAAAATATAAATATAATTTATAT 513

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
              ||  |  |  |||   | || | |  ||| ||   |  ||   ||||  ||| |||
Db        512 AAATTATATATCAATATAAAATATAAATATAATTTATATAAATTATATATCAATATAAAA 453

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187
              |  |  ||   | | ||| ||   | |   | ||  || |||  ||||   | |
Db        452 TATAAATATAATTTATATAAATTATATATCAATATAAAATATAAATATAATTTATATAAA 393

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247
              |  || || ||||||||  | |  | |  |   |   || | | | || |  |
Db        392 TAAAATTATATAAATATAATTTATATAAATAAAATTATATAAATATAATTTATATAAATA 333

Qy        248 ATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAA 307
              | |||||  |  |  | ||||  |   |  || | ||  |  ||  |||||| | |
Db        332 AAATTATATGAATATAATTTATATAAATATAATTTATATAAATAAAATTATATGAATATA 273

Qy        308 ATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGA 367
              || |  | ||   ||  ||| |  | | |       | || || ||  | |
Db        272 ATTTATATAAATAAAATTATATAAATATAATTTATATAAATAAAATTATGAATATAATTT 213

Qy        368 AAATTGATTCATTGGATA 385
              | ||  ||  | |  |||
Db        212 ATATGAATAGAATTTATA 195


RESULT 10
BX248097
LOCUS       BX248097               174835 bp    DNA     linear   HTG 03-APR-2003
DEFINITION  Danio rerio clone CH211-232M8, *** SEQUENCING IN PROGRESS ***, 4
            unordered pieces.
ACCESSION   BX248097
VERSION     BX248097.3  GI:29538764
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 174835)
  AUTHORS   Mclaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2003 this sequence version replaced gi:28200814.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
```

```
            ---------- Project Information
            Center project name: zC232M8
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173639 bases at least Q40
            Consensus quality: 173869 bases at least Q30
            Consensus quality: 174126 bases at least Q20
            Insert size: 174535; sum-of-contigs
            Insert size: 177115; 1.4% error; agarose-fp
            Quality coverage: 12.09x in Q20 bases; sum-of-contigs Quality
            coverage: 12.10x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    46372: contig of 46372 bp in length
            *    46373    46472: gap of 100 bp
            *    46473    95245: contig of 48773 bp in length
            *    95246    95345: gap of 100 bp
            *    95346   170226: contig of 74881 bp in length
            *   170227   170326: gap of 100 bp
            *   170327   174835: contig of 4509 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174835
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-232M8"
                     /clone_lib="CHORI-211"
     misc_feature    1. .46372
                     /note="assembly_fragment:01305
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    46473. .95245
                     /note="assembly_fragment:01589
                     fragment_chain:1"
     misc_feature    95346. .170226
                     /note="assembly_fragment:00787
                     fragment_chain:1"
     misc_feature    170327. .174835
                     /note="assembly_fragment:00604
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    55674 a  30596 c  31707 g  56558 t    300 others
ORIGIN

  Query Match             15.4%;  Score 68.8;  DB 2;  Length 174835;
  Best Local Similarity   48.9%;  Pred. No. 0.12;
  Matches  215;  Conservative    0;  Mismatches  222;  Indels    3;  Gaps    1;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              | ||||| ||||| |||       || || || |   | |   ||||  ||    || ||
Db      94516 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 94575

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
              | |    || ||| |   |   |   || || | ||| || ||| | || |      || |
Db      94576 ATATATACATATATATATATATTAATTAAAAATTAAAATGAACAAAAAAACAAAATAA 94635

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              ||    ||| |||  | ||| ||  |       | |   ||  | |   |  ||  | |
Db      94636 AATATAATTGCAATAAATTTGAAATAAAATAAAAAGGAAAATAAAATAATAAACTATATA 94695

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              |    ||||| || || | ||||| ||   |  |       | | | || | | |     |
Db      94696 AAAATATATATATGAAAAAAATAAATCGAATAAATAAATAATTTAATTTAGTTAAAATAA 94755

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              | |||  ||   | |||   | | |  |     | |||||  | |  || |  |   ||
Db      94756 AAATAAAATAAAATAAAAACCAAATAAAATACAAAATAAAATAATAAAATACAAAATAAT 94815

Qy        306 AAATA---TTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAAC 362
              | |||   | ||| |   || |||||| ||  |   |||  || |  ||  ||  |||  ||
Db      94816 ATATAGAGTAAAATAAAAAATAAATAAAATACAAAATATATATATTTAAATAAAAATAAA 94875

Qy        363 TTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAG 422
              |   |||       |||  | |      | | | || ||  |  ||   || ||  ||
Db      94876 ATAATAAATAACAAAATTTAAAATAAAAATAATAAAATTAAATTAAATAATAAAAAGTAA 94935

Qy        423 TGAAAACCCCGATGATATGT 442
              | | ||     || | | ||
Db      94936 TTACAAATAAAATAAAAAGT 94955


RESULT 11
AL929250/c
LOCUS       AL929250               94434 bp    DNA     linear   VRT 22-JAN-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-137A2, complete sequence.
ACCESSION   AL929250
VERSION     AL929250.7  GI:27848021
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 94434)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 22, 2003 this sequence version replaced gi:27803921.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU).  For
            further information see http://www/Projects/D_rerio/fishmask.shtml
            CH211-137A2 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .94434
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
```

```
                  /clone="CH211-137A2"
                  /clone_lib="CHORI-211"
BASE COUNT    28606 a  17026 c  17319 g  31483 t
ORIGIN

  Query Match             15.1%;  Score 67.6;  DB 5;  Length 94434;
  Best Local Similarity   48.8%;  Pred. No. 0.21;
  Matches  211;  Conservative    0;  Mismatches  219;  Indels    2;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
Db      30585 ATAATAATATAGATAATAAAGAATATAGATAATAATAATAATAATAATAATGATGATAAT 30526

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
Db      30525 AATAGTAATAATAATGATAATAATAATAATAATTATAATATTAATAACAATAATAATAAT 30466

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187
Db      30465 ATGGATAATAAAGATAAT--ATAGTTAATAATAATAATAATAATATTAATAATGATGATA 30408

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247
Db      30407 ATAATAATAATGATGATAATAATATTATTATTATTAATATTAATAATAATATAGACAATA 30348

Qy        248 ATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAA 307
Db      30347 AAGATAATATAGTTAATAATAATAATAATGATGATAATAAAGATAATAATAATAATAATA 30288

Qy        308 ATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGA 367
Db      30287 ATAGTATTAATGATAATAATAATGATAATAATAATAATATTAATAATAATAATAATGATA 30228

Qy        368 AAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAAA 427
Db      30227 ATGATGGTAATTATAATAATAATAGTAATAATAATGATAATAATGATAATGATAATAATA 30168

Qy        428 ACCCCGATGATA 439
Db      30167 ATATTAATAATA 30156

RESULT 12
AC125567/c
LOCUS       AC125567              170627 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC125567
VERSION     AC125567.5  GI:24817949
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 170627)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170627)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 170627)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096537.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDDV
                Center clone name: CH230-9H22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162619 bases at least Q40
                Consensus quality: 164392 bases at least Q30
                Consensus quality: 165269 bases at least Q20
                Estimated insert size: 169231; sum-of-contigs estimation
```

```
             Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
           --------------
           * NOTE: Estimated insert size may differ from sequence length
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 4 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1   109158: contig of 109158 bp in length
           *   109159   109258: gap of unknown length
           *   109259   110280: contig of 1022 bp in length
           *   110281   110380: gap of unknown length
           *   110381   111540: contig of 1160 bp in length
           *   111541   111640: gap of unknown length
           *   111641   170627: contig of 58987 bp in length.
FEATURES             Location/Qualifiers
     source          1..170627
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9H22"
     misc_feature    1..1284
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3636..4924
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH341988"
BASE COUNT    51711 a  30081 c  30017 g  54086 t   4732 others
ORIGIN

  Query Match              15.0%;  Score 67.2;  DB 2;  Length 170627;
  Best Local Similarity    49.4%;  Pred. No. 0.2;
  Matches  174;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy          9 TATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGA 68

Db      28794 TATTAAAAAAATAAAAATATATAATATATAAAAAAAAAAAAAAAAAATAAAAATATAAAA 28735

Qy         69 AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAAC 128

Db      28734 AAAAAAAAAAAAAAATAAAAAAATTAAAAAAAAAATAATATTTAAAATTAAAAAAAAAAA 28675

Qy        129 TGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATC 188

Db      28674 TAAAAAAAAAAAAAATATAAAAAATTTAAAAAAAAAAATAAAAATAAAAAAAAAAAATAAA 28615

Qy        189 CCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGA 248

Db      28614 ATAAAAAAATAATTAATAAAAAAAATATTAATATAAAAAAAATAAAAAAAAATTATAAAA 28555

Qy        249 TATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAA 308

Db      28554 AATAAAAATAAAAATATATAAATAAAAAAATAAAATATAAATAATAATTTAAATAAAAAA 28495

Qy        309 TATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Db      28494 ATTATAAAAAAAAAAAAATAAAATAAAAAAAAAAAATAAAAAAATTAAAAAA 28443


RESULT 13
PFMAL7P1_04/c
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL7P1 Accession AL844506
     Fragment Name       Begin      End
     PFMAL7P1_00             1      110000
     PFMAL7P1_01        100001      210000
     PFMAL7P1_02        200001      310000
     PFMAL7P1_03        300001      410000
     PFMAL7P1_04        400001      510000
     PFMAL7P1_05        500001      610000
     PFMAL7P1_06        600001      710000
     PFMAL7P1_07        700001      810000
     PFMAL7P1_08        800001      910000
     PFMAL7P1_09        900001     1010000
     PFMAL7P1_10       1000001     1110000
     PFMAL7P1_11       1100001     1210000
     PFMAL7P1_12       1200001     1310000
     PFMAL7P1_13       1300001     1351552
Continuation (5 of 14) of PFMAL7P1 from base 400001 (AL844506 Plasmodium falciparum 3D7

  Query Match              14.9%;  Score 66.4;  DB 2;  Length 110000;
  Best Local Similarity    47.2%;  Pred. No. 0.3;
  Matches  202;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Db      19324 ATTTTATATATATATATTTATATTGATAAAGAAAAAAATGAAAAAAAAGAAAAAAAAAAA 19265

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Db      19264 TTTATTAAAATAAAAAAAAAAAGAGAAAGATTTTAAAAATAATAAAAATTATAATAAAAA 19205

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Db      19204 TATAAATTTTGATAAAATAAAAAATGAAAAATTTTATCAAAAAAAATTAAAAACAAATTA 19145

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Db      19144 TATTAATAAAAATAATGATTATAAAAAAAAATTATTAGAAATAAAATAAAAACAAAAAAA 19085

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Db      19084 GAAAAATAAAATGTTAAAAAAAAAATATATATCATAAAATAAAAAGAAAAGGGAAAAAAT 19025

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Db      19024 GTTAAAAAAAAAAATAAAATTAAAAAAAAAAAAAAAAAAAAAAATTAAAAAATAAAATA 18965

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Db      18964 AAATAAAAAAAAATTAAATAAAATAAAATTAAAAAAAAAAAAAAATTAAAAAATTTATTT 18905

Qy        421 AGTGAAAA 428

Db      18904 AAATAAAA 18897


RESULT 14
AC117140/c
LOCUS       AC117140              286208 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-365F7, WORKING DRAFT SEQUENCE, 9
            unordered pieces.
ACCESSION   AC117140
VERSION     AC117140.4  GI:24635509
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 286208)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
```

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 286208)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 286208)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 6, 2002 this sequence version replaced gi:23097379.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GUCU
Center clone name: CH230-365F7
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 125022 bases at least Q40
Consensus quality: 128107 bases at least Q30
Consensus quality: 129968 bases at least Q20
Estimated insert size: 130783; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47440: contig of 47440 bp in length
* 47441 47540: gap of unknown length
* 47541 230268: contig of 182728 bp in length
* 230269 230368: gap of unknown length
* 230369 254735: contig of 24367 bp in length
* 254736 254835: gap of unknown length
* 254836 275940: contig of 21105 bp in length
* 275941 276040: gap of unknown length
* 276041 277252: contig of 1212 bp in length
* 277253 277352: gap of unknown length
* 277353 278493: contig of 1141 bp in length
* 278494 278593: gap of unknown length
* 278594 282220: contig of 3627 bp in length
* 282221 282320: gap of unknown length
* 282321 284049: contig of 1729 bp in length
* 284050 284149: gap of unknown length
* 284150 286208: contig of 2059 bp in length.

FEATURES Location/Qualifiers
source 1. .286208
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-365F7"
misc_feature 6333. .7149
/note="clone_boundary
clone_end:T7
site:
end_sequence:BZ116846"
misc_feature 6368. .7302
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ116847"
misc_feature 47541. .48868
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 230369. .232081
/note="wgs_end_extension
clone_end:Sp6"

BASE COUNT 43136 a 24292 c 24002 g 39383 t 155395 others
ORIGIN

Query Match 14.9%; Score 66.4; DB 2; Length 286208;
Best Local Similarity 48.6%; Pred. No. 0.23;
Matches 207; Conservative 0; Mismatches 218; Indels 1; Gaps 1;

```
Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              |||  | | | | | ||  |||||||   | || || || ||    || ||  ||| ||
Db     229766 AAATTAAAAAAAAAAAAAAATATTAAAAAAATAATAAAAAAAAACATAAATATAATTAAA 229707

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
                ||   |||||     || | || || || |   | ||   ||| || |     | |
Db     229706 TAAAAAATAAATAATTAAAAAATAAAATTAAAAAAAATAAAATAAAAAAAAAAAAAATA 229647
```

```
Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              |  |  || | ||  |     ||  |      ||  || || ||  |   |||  |
Db     229646 ATATAAATATTAAAAAAAAAAAAAAAAAAAAATAATTTAAAAAAAAATATAAAAAATAAAA 229587

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              |         | || ||||| | |||  |||    |        ||   ||| |||||
Db     229586 AAAATGATAAAATAAATATTACAACAATTCAAAAAAAAAAAAAAAATTTTAAATAAAAAAAT 229527

Qy        246 AGATATTATGGCACAAATTTCT-ATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAA 304
              | | ||||    |    | |   ||||  ||  || |  || ||  | || |  |  | |
Db     229526 AAAAATTAGAAAATTTTTATAAAATAAAAAAAAGAAAATAAAAAAATAAAAAAAAATATA 229467

Qy        305 AAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTT 364
              ||||||  |||||   || ||| |||  || | | | |   |  |    ||| ||| |
Db     229466 AAAATAAAAAAAAATTAATAAAAAAAAAAAAAATAAAAAGAAAAAATTAAAAAAAAATA 229407

Qy        365 GGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTG 424
                ||| | || ||  | |  |  |    |  || ||     ||  ||| ||  |
Db     229406 ATCAAAATAATACAAAGAAATACNTTAACTATAAAATAAAAACAAAATTAAAAACNAAAA 229347

Qy        425 AAAACC 430
              |||| |
Db     229346 AAAATC 229341
```

RESULT 15
AF437291

```
LOCUS       AF437291               25753 bp    DNA     circular PLN 31-MAY-2002
DEFINITION  Saccharomyces castellii strain NRRL Y-12630 mitochondrion, complete
            genome.
ACCESSION   AF437291
VERSION     AF437291.1  GI:21105284
KEYWORDS    .
SOURCE      mitochondrion Saccharomyces castellii
  ORGANISM  Saccharomyces castellii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 25753)
  AUTHORS   Petersen,R.F., Langkjaer,R.B., Hvidtfeldt,J., Gartner,J.,
            Palmen,W., Ussery,D.W. and Piskur,J.
  TITLE     Inheritance and organisation of the mitochondrial genome differ
            between two Saccharomyces yeasts
  JOURNAL   J. Mol. Biol. 318 (3), 627-636 (2002)
  MEDLINE   22050826
   PUBMED   12054811
REFERENCE   2  (bases 1 to 25753)
  AUTHORS   Petersen,R.F., Langkjaer,R.B., Hvidtfeldt,J., Gartner,J.,
            Palmen,W., Ussery,D.W. and Piskur,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) BioCentrum-DTU, Section of Molecular
            Microbiology, Technical University of Denmark, Building 301, Lyngby
            DK-2800, Denmark
FEATURES             Location/Qualifiers
     source          1. .25753
                     /organism="Saccharomyces castellii"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /strain="NRRL Y-12630"
                     /db_xref="taxon:27288"
     tRNA            224. .297
                     /product="tRNA-Val"
                     /note="codon recognized: GUA"
                     /evidence=not_experimental
     gene            507. .1316
                     /gene="COX3"
                     /evidence=not_experimental
     CDS             507. .1316
                     /gene="COX3"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="cytochrome c oxidase subunit III"
                     /protein_id="AAM34588.1"
                     /db_xref="GI:21105285"
                     /translation="MTHLERSKYQQFPFHMVMPSPWPLMVSLSILSLALSLAFTMHGF
                     INNMNILYLSLFVLTLSSTFWFRDVISEATYLGDHTIAVRKGINLGFLLFVLSEVLIF
                     ASLFWAYFHSAMSPDIILGAQWPPVGIEAVQPTELPLLNTIILLSSGATITYSHHALI
                     NKDRKNALMGLLITTILIVIFVSCQYIEYINAMFTISDGVYGTVFYAGTGLHFLHMVM
                     LAVMLMINYWRMRNYHLTSGHHVGYETTIIYCHVLDVIWLFLYIIFYWWAV"
     tRNA            1936. .2017
                     /product="tRNA-Leu"
                     /note="codon recognized: UUA"
                     /evidence=not_experimental
     tRNA            2251. .2323
                     /product="tRNA-Gln"
                     /note="codon recognized: CAA"
                     /evidence=not_experimental
     tRNA            2420. .2491
                     /product="tRNA-Lys"
                     /note="codon recognized: AAA"
                     /evidence=not_experimental
     tRNA            2501. .2573
                     /product="tRNA-Arg"
                     /note="codon recognized: AGA"
                     /evidence=not_experimental
     tRNA            2592. .2663
                     /product="tRNA-Gly"
                     /note="codon recognized: GGA"
                     /evidence=not_experimental
     tRNA            2713. .2784
                     /product="tRNA-Asp"
                     /note="codon recognized: GAC"
                     /evidence=not_experimental
     tRNA            2926. .3010
                     /product="tRNA-Ser"
                     /note="codon recognized: AGC"
                     /evidence=not_experimental
     tRNA            3153. .3223
                     /product="tRNA-His"
                     /note="codon recognized: CAC"
                     /evidence=not_experimental
     tRNA            3292. .3364
                     /product="tRNA-Met"
                     /note="codon recognized: AUG"
                     /evidence=not_experimental
     gene            3857. .4837
                     /gene="VAR1"
                     /evidence=not_experimental
     CDS             3857. .4837
                     /gene="VAR1"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="mitochondrial ribosomal protein"
                     /protein_id="AAM34589.1"
                     /db_xref="GI:21105286"
                     /translation="MKLNILKLLNNNNNKNMLLKNMLLKMNINNNNMNTNNNNMKYMN
                     NKLQSINNMNNWSLQMYNYNKNNELNMMMMNMNMNKLLYKLMNMIMINNMTMNVNNKN
                     NNNIIMMKPMYQYMMNKLNIKFYYYINNNNMNTNYYMNMLYKLMNTLNNNNNMYMNNM
                     NNIMSMYINKNINIEMIKLNYVYNNKDIMNKYLSTMDIDTLNNNSTMNLLNNMMTKMN
                     NNNIIMNYMNNMNNMRNNKYINNMSSNYIMNMLMYKYLTGWTILLKGRLNKNMTRTNK
                     YILYNGSNKMNMYMKNNYKLNYISNNHFINNINNVNKNGKYNIKVKLSYI"
     tRNA            4940. .5011
                     /product="tRNA-Glu"
                     /note="codon recognized: GAA"
                     /evidence=not_experimental
     gene            5489. .6661
                     /gene="COB"
                     /evidence=not_experimental
     CDS             5489. .6661
                     /gene="COB"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="apocytochrome b"
```

```
                     /protein_id="AAM34590.1"
                     /db_xref="GI:21105287"
                     /translation="MTFRKSNVYLNLVNSYIIDSPQPSSINYWWNMGSLLGLCLVIQI
                     LTGIFMAMHYSSNIELAFSSVEHIMRDVQGGWFLRYAHANGASFFFICMYIHMGKALY
                     YGSYRSPRVLLWTIGVIIFILTMATAFLGYCCVYGQMSHWGATVITNLFSAIPFIGKD
                     IVLWLWGGFAVSNPTIQRFFALHYLFPFVIAAVVIMHMMALHIHGSSNPLGITGNMDR
                     LPMHGYFVFKDLITVFVFLIVFSLFVFFSPNTMGHPDNYIPGNPMVTPASIVPEWYLL
                     PFYAILRSIPDKLMGVITMFSAILVLLVLPFTDRSVVRGNSFKVLSKLFFFLFVFNFV
                     LLGQIGAVHVEVPYILMGQISTFLYFAYFLVFIPIISTIENILFYVGSRNNTDDLK"
tRNA                 8530. .8603
                     /product="tRNA-Met"
                     /note="codon recognized: AUG"
                     /evidence=not_experimental
gene                 8604. .9116
                     /gene="RPM1"
                     /evidence=not_experimental
misc_RNA             8604. .9116
                     /gene="RPM1"
                     /product="RNase P RNA"
                     /evidence=not_experimental
tRNA                 9219. .9290
                     /product="tRNA-Pro"
                     /note="codon recognized: CCA"
                     /evidence=not_experimental
gene                 9644. .12858
                     /gene="COX1"
                     /evidence=not_experimental
CDS                  join(9644. .9883,10371. .10850,11974. .12858)
                     /gene="COX1"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="cytochrome c oxidase subunit I"
                     /protein_id="AAM34591.1"
                     /db_xref="GI:21105288"
                     /translation="MLQRWLYSTNAKDIAVLYFLLALFSGMAGTAMSLIIRLELAAPG
                     QQYLHGNNQLFNVLVVGHAILMIFFMVMPALIGGFGNYMLPLMIGATDTAFPRINNIG
                     FWLLPMGLVCLVTSTLVESGAGTGWTVYPPLASIQAHSGPSVDLAIFALHMTSISSLL
                     GAINFIVTTLNMRTNGMTMHKLPLFVWAILITAVLLLLTLPVLSAGVTMLLLDRNFNT
                     SFFEVAGGGDPILYEHLFWFFGHPEVYILIIPGFGIISHVVSTYSKKPVFGEISMVYA
                     MASIGLLGLLVWSHHMYIVGLDADTRAYFTSATMIIAIPTGIKIFSWLATIYGGSIRL
                     ATPMLFAIAFLFLFTMGGLTGVALANASLDVAFHDTYYVVGHFHYVLSMGAVFSLLAG
                     YYYWSPQILGLYYNEKLSQIQFWLIFVGANMVFLPMHFLGINGMPRRIPDYPDAFAGW
                     NYISSIGAVVSMISLLLFVYILYDQLVNGLTNKINNKAVVFYKTPDFMESNMVFNTNT
                     VKSSSLEFLLTSPPAVHSFNTPAVQS"
CDS                  join(9644. .9883,10371. .11927)
                     /note="LAGLIDADG type group IB intron-encoded
                     endonuclease"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="I-SceII DNA endonuclease-like protein"
                     /protein_id="AAM34596.1"
                     /db_xref="GI:21105293"
                     /translation="MLQRWLYSTNAKDIAVLYFLLALFSGMAGTAMSLIIRLELAAPG
                     QQYLHGNNQLFNVLVVGHAILMIFFMVMPALIGGFGNYMLPLMIGATDTAFPRINNIG
                     FWLLPMGLVCLVTSTLVESGAGTGWTVYPPLASIQAHSGPSVDLAIFALHMTSISSLL
                     GAINFIVTTLNMRTNGMTMHKLPLFVWAILITAVLLLLTLPVLSAGVTMLLLDRNFNT
                     SFFEVAGGGDPILYEHLFWFFGTTIGPLMLYMLMNIPKCWKLLKFMITNMMSTFMALL
                     VKMFMNMYNQQDKIRYNLNYIKYYINLLMQFLFNKNYTKRSSETTWDMLNMSKSETVN
                     QSSGSMMSCADLLMSTTINNKMLNKEDIKFNQWLAGLIDGDGYFGITQNKYTSCEITM
                     ALEDEMALKLIQNKLGGSIKLRSGVRAVRYRLTNKKGMINLINTVNGNIRNTKRLLQF
                     YQVCNILDIMPINPIKLDYDNAWFIGFFDADGTITYSFKGKYPLLTISMTNKYLVDVQ
                     PIKDILGGYIYPDKAKNGYYKWSIQSKQDVLDFINNYVKSNPSRTTKMNRLMLSQEFY
                     KLLEMRAYNKNYNTAMYKAWLNFENKWKSNNI"
gene                 13285. .13431
                     /gene="ATP8"
                     /evidence=not_experimental
CDS                  13285. .13431
                     /gene="ATP8"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=3
                     /product="ATP synthase subunit 8"
                     /protein_id="AAM34592.1"
                     /db_xref="GI:21105289"
                     /translation="MPQLLPFYFLNTLTYGFILIMILLVLFSQYFLPMILRLYVSRLF
                     ISKL"
gene                 13837. .14616
                     /gene="ATP6"
                     /evidence=not_experimental
CDS                  13837. .14616
```

```
Query Match          14.7%;  Score 65.8;  DB 8;  Length 25753;
Best Local Similarity  51.0%;  Pred. No. 0.53;
Matches  180;  Conservative    0;  Mismatches  172;  Indels    1;  Gaps    1;

Qy        20 ATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTCAAATAC 79
             || || ||| |  |||| ||  |    |||| |||||  || ||||     |  | ||
Db      3758 ATTATATATAATCTTTATAATTATTTTTATATAAATTAATATATAAATGGATTAGATTAT 3817

Qy        80 AAGGAACA-AAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTTCAA 138
             ||   | | |  ||  ||||  || ||  || | ||||     |   || |   ||  ||
Db      3818 AATATATATATTGAATAATACATATAATAAATAAAATAAATGAAATTAAATATTTTAAAA 3877

Qy       139 CGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATAGAT 198
             |||     | | | |||   |   |||| | |||  |   ||||   | ||| ||
Db      3878 TTATTAAATAATAATAATAATAAAAACATATTATTAAAAAATATATTATTAAAAATAAAT 3937

Qy       199 CAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATGGCA 258
             |||||||  |       |||  ||| ||  | || |    | ||| ||    |
Db      3938 ATTAATAATAATAATATAAATCTTAATAATAATAATATAAAATATATAAATAATAAATTA 3997

Qy       259 CAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAAAAG 318
             |||     ||  | |       ||  | | |  ||   ||| |  | ||||  || ||
Db      3998 CAATCAATTAATAATATAAATAATTGATCATTACAAATATATAATTATAATAAAAATAAT 4057

Qy       319 GCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAAT 371
             | |  ||||| | ||| | | |||  ||  |||| |||  |   |  ||| |
Db      4058 GAATTAAATATAATAATAATAATAAATATAATAAATAAATTATTATATAAATT 4110
```

Search completed: February 16, 2004, 12:08:36
Job time : 2886 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:28:37 ; Search time 65 Seconds
(without alignments)
1004.995 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980054/runat_13022004_163841_26344/app_query.fasta_1.327
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054_@CGN_1_1_56_@runat_13022004_163841_26344 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	91.5	11.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
2	86	11.1	7100	4	US-09-308-375-1	Sequence 1, Appli
c 3	84	10.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
4	83.5	10.8	3258	2	US-08-729-955A-2	Sequence 2, Appli
5	82	10.6	2271	4	US-09-134-001C-716	Sequence 716, App
6	81	10.5	1563	4	US-09-508-370A-11	Sequence 11, Appl
7	80	10.4	6773	4	US-09-166-350-27	Sequence 27, Appl
8	80	10.4	1230025	4	US-09-198-452A-1	Sequence 1, Appli
9	78	10.1	417	4	US-09-252-991A-6512	Sequence 6512, Ap
c 10	78	10.1	747	4	US-09-252-991A-6539	Sequence 6539, Ap
11	78	10.1	1521	4	US-09-252-991A-6495	Sequence 6495, Ap
12	78	10.1	3312	3	US-08-923-992A-3	Sequence 3, Appli
13	77.5	10.0	1488	3	US-08-867-611-9	Sequence 9, Appli
14	77.5	10.0	1488	5	PCT-US92-06965A-14	Sequence 14, Appl
15	77.5	10.0	1497	1	US-08-500-222-1	Sequence 1, Appli
16	77.5	10.0	1497	1	US-08-500-125-1	Sequence 1, Appli
17	77.5	10.0	1497	2	US-07-779-704B-1	Sequence 1, Appli
18	77	10.0	1275	3	US-08-867-611-21	Sequence 21, Appl
19	77	10.0	1275	5	PCT-US92-06965A-26	Sequence 26, Appl
20	77	10.0	1494	4	US-09-328-352-3160	Sequence 3160, Ap
21	77	10.0	1590	4	US-08-887-534A-73	Sequence 73, Appl
22	77	10.0	1590	4	US-09-527-431-73	Sequence 73, Appl
23	77	10.0	1851	3	US-08-867-611-29	Sequence 29, Appl
24	77	10.0	1851	5	PCT-US92-06965A-34	Sequence 34, Appl
25	76.5	9.9	1491	3	US-08-926-842B-35	Sequence 35, Appl
26	76.5	9.9	2433	4	US-09-540-824-24	Sequence 24, Appl
27	76.5	9.9	10917	3	US-08-926-842B-11	Sequence 11, Appl
28	76	9.8	2617	1	US-08-430-024-1	Sequence 1, Appli
29	76	9.8	2617	1	US-08-782-009-1	Sequence 1, Appli
30	76	9.8	2617	3	US-09-017-302-1	Sequence 1, Appli
31	76	9.8	3730	1	US-08-242-932-8	Sequence 8, Appli
32	76	9.8	3730	1	US-08-714-481-8	Sequence 8, Appli
33	76	9.8	3730	5	PCT-US95-06111-8	Sequence 8, Appli
34	76	9.8	4200	1	US-08-242-932-1	Sequence 1, Appli
35	76	9.8	4200	1	US-08-714-481-1	Sequence 1, Appli
36	76	9.8	4200	3	US-08-923-992A-1	Sequence 1, Appli
37	76	9.8	4200	5	PCT-US95-06111-1	Sequence 1, Appli
38	75.5	9.8	777	4	US-09-107-532A-835	Sequence 835, App
39	75.5	9.8	993	4	US-09-134-001C-974	Sequence 974, App
40	75.5	9.8	2043	4	US-09-601-198-181	Sequence 181, App
41	75.5	9.8	2404	1	US-08-484-105-7	Sequence 7, Appli
42	75.5	9.8	2404	1	US-08-484-106-7	Sequence 7, Appli
43	75.5	9.8	3636	4	US-09-134-001C-1983	Sequence 1983, Ap
44	75	9.7	3033	4	US-09-134-001C-2341	Sequence 2341, Ap
45	74.5	9.7	789	1	US-08-160-670A-31	Sequence 31, Appl

ALIGNMENTS

```
RESULT 1
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
```

; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Alignment Scores:

Pred. No.:	451	Length:	1664976
Score:	91.50	Matches:	43

```
Percent Similarity:     44.25%          Conservative:  34
Best Local Similarity:  24.71%          Mismatches:    62
Query Match:            11.85%          Indels:        35
DB:                     4               Gaps:          7

US-09-980-054A-12 (1-148) x US-08-916-421B-1 (1-1664976)

Qy          3 AspIleAsp---IleAspAsnValLeuAsn---------------------------- 11
              |||||||||   :::|||   :::||||||
Db     199992 GATATTGATAATTTAGATACTCTATTAAATAAAATAAAAGATGAAATTGAAAGAGTGGAA 200051

Qy         12 ------------LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGln 27
                        |||   ||||||          |||      :::      :::|||:::
Db     200052 ACTATTAAAGATTTGCTTGAAGAACTTAAAAATCTAAATGAAGAGATTGAAAAAATTGAA 200111

Qy         28 GlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArg 47
                 |||              |||   |||||||||                 :::::::::
Db     200112 AAATATAAAAGAATATGTGAAGAGTGCAAAGAATAC--------------TATGAGAAG 200156

Qy         48 PheLeu---IleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAsp 66
              :::|||   :::              |||     |||     |||    :::|||
Db     200157 TATTTAGAATTAGAAGAAAAGGCTGTAGAATACAATAAACTAACTTTAGAGTATATAACA 200216

Qy         67 GlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAla 86
                        |||:::      ::::::|||:::|||||      :::
Db     200217 TTGCTTCAGGAGAAAAAATCCATTGAAAAAAATATTAACGATTTAGAAACAAGAATTAAT 200276

Qy         87 GlnIle------------------------------SerIleThrAsnGlyAspLysGlu 96
              ::::::                            ||||||   |||      ||||||
Db     200277 AAACTTTTAGAAGAAACAAAAAATATTGACATTGAAAGTATTGAAAATTCATTAAAAGAG 200336

Qy         97 ValGluAspTyrGluLys---------AsnIleLysLysAlaArgAsnLysLeu---Arg 112
              :::|||:::   :::|||         |||:::::::|||   :::   :::|||   :::
Db     200337 ATAGAGGAAAAAAAGAAAGTTCTTGAAAATCTACAAAAAGAAAAGATAGAACTAAACAAA 200396

Qy        113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
                 :::      |||      |||      ::::::      :::   :::   :::|||
Db     200397 AAACTTGGGGAAATTAACAGTGAAATTAAAAGGCTGAAAAAAATTTTAGATGAACTTAAA 200456

Qy        133 GluValGlyGlyThrLeuGlnValSerGluAsnProAspAsp 146
              ||||||   |||      :::   :::   |||   |||
Db     200457 GAAGTTGAAGGAAAATGCCCACTATGTAAAACACCAATAGAT 200498

RESULT 2
US-09-308-375-1
; Sequence 1, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
;  APPLICANT: Genencor International, Inc.
;  TITLE OF INVENTION: Proteases From Gram-Positive Organisms
;  FILE REFERENCE: GC394-PCT
;  CURRENT APPLICATION NUMBER: US/09/308,375
;  CURRENT FILING DATE: 1999-05-14
;  EARLIER APPLICATION NUMBER: EP9719636.4
;  EARLIER FILING DATE: 1997-09-15
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 7100
;   TYPE: DNA
;   ORGANISM: Bacillius subtilis
US-09-308-375-1

Alignment Scores:
Pred. No.:              1.23            Length:        7100
Score:                  86.00           Matches:       35
Percent Similarity:     44.23%          Conservative:  34
Best Local Similarity:  22.44%          Mismatches:    61
Query Match:            11.14%          Indels:        26
DB:                     4               Gaps:          6

US-09-980-054A-12 (1-148) x US-09-308-375-1 (1-7100)
```

```
Qy          5 AspIleAspAsnValLeuAsnLeu---GluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              |||:::||||||   :::::::::   ||||||::::::   ::::::   :::
Db       5677 GATGTCGATAACAAAATTTCAATGACTGAAGAAGATGAAGACAAGGTTAAATACTATAGC 5736

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
                ||||||:::      :::   |||      |||   |||:::|||      |||
Db       5737 AAGCAAATAAAGCTCATTCAACAACAACAAAAGGAAGCGAAGAAATACATTAAGCAGCTT 5796

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyr-----------------IleGlnGluLeu 57
                 |||:::            |||:::                   |||   |||
Db       5797 GAAGAACAAAAGAAAGCTGCGAAAGGTTTCCCTGACATCCAGGAACAGATCACTGAAGAA 5856

Qy         58 MetLysPheTrpLeuSer--------------HisIleAspGlnTyrAsnAsnSerSer 72
              |||:::   |||                   :::::::::   ||||||
Db       5857 ATGCAAAACTGGAAAGATAAACAGAAAGATTTTAACCTTGAGCTTTATAACACCAAGAAG 5916

Qy         73 SerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsn 92
              |||:::::::::      :::|||  |||   :::      |||
Db       5917 TCGATCAAGGATATCTATAAATCATTGGCTGATGAAGTTGTATCCATC------------ 5964

Qy         93 GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsn---LysLeu 111
                                  |||::::::      :::|||   |||:::   :::|||
Db       5965 --------------------TACAAAGAGATGTACGAAAAAATGCGTGATATTGAGTTA 6003

Qy        112 ArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuVal 131
                              |||::::::         ||||||   :::|||
Db       6004 GAAGCGCATCAGAAAGCGACTCAAGACTTG------ATCGATGAGATAGACAAGACTGAT 6057

Qy        132 LysGluValGlyGlyThrLeuGlnValSerGluAsnProAspAspMet 147
                |||               ::::::   |||      |||   :::
Db       6058 GACGAGGCTAAATTTCAAAAAGAATTAAAAGAAAGACAAGACAGTATT 6105

RESULT 3
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Alignment Scores:

Pred. No.:	4.07e+03	Length:	1664976
Score:	84.00	Matches:	32
Percent Similarity:	42.50%	Conservative:	19
Best Local Similarity:	26.67%	Mismatches:	45
Query Match:	10.88%	Indels:	24
DB:	4	Gaps:	4

```
US-09-980-054A-12 (1-148) x US-08-916-421B-1 (1-1664976)

Qy          7 AspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPhe---------- 21
Db    1122980 GATAAAAAATTTGAGTTAATTAAAGAATACGATGAGAAAGGTTGGATATGTAGTCTTAAA 1122921

Qy         22 LysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyr 41
Db    1122920 AAAGCTGGAAAACTACAACCTTACAAAATCTACATAATTGATGACAAATATTACTCTAAA 1122861

Qy         42 GlnThrGlyPheGln------ArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLys 59
Db    1122860 GAAACTGGGATAATAATTTAACAAACTTTGATTTTGGGTGAGATTATGGAAGTTTTAGCC 1122801

Qy         60 PheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsn 79
Db    1122800 TTT-------------------------AAAAATCAATCCCTTATTGATCATGTTAAC 1122768

Qy         80 AsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAsp 99
Db    1122767 GATATGGTCAAATATTGGGAGAGAATAAAATATAGA---------------------- 1122732

Qy        100 TyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLys 119
Db    1122731 TACTTAAAAACTATAAAGAGAGCTTTAGAGGCTTTAAACATAAAGTTAGATATCGAAAAA 1122672

RESULT 4
US-08-729-955A-2
; Sequence 2, Application US/08729955A
; Patent No. 5932417
;  GENERAL INFORMATION:
;    APPLICANT:  Birnbaumer, Lutz
;    APPLICANT:  Zhu, Xi
;    TITLE OF INVENTION:  Method And Compounds For Controlling
;    TITLE OF INVENTION:  Capacitative Calcium Ion Entry Into Mammalian Cells
;    TITLE OF INVENTION:  Essential for Agonist-Activated Capacitative Ca2+
;    TITLE OF INVENTION:  Entry
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Poms, Smith, Lande & Rose
;      STREET:  2029 Century Park East, Suite 3800
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,955A
;      FILING DATE:  October 15, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/025,111
;      FILING DATE:  August 29, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oldenkamp, David J.
;      REGISTRATION NUMBER:  29,421
;      REFERENCE/DOCKET NUMBER:  120186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (310) 788-5000
;      TELEFAX:  (310) 277-1297
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3258 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:
;      INDIVIDUAL ISOLATE:  Mtrp4
US-08-729-955A-2

Alignment Scores:
Pred. No.:                0.876          Length:        3258
Score:                    83.50          Matches:       37
Percent Similarity:       45.21%         Conservative:  29
Best Local Similarity:    25.34%         Mismatches:    49
Query Match:              10.82%         Indels:        31
DB:                       2              Gaps:          6

US-09-980-054A-12 (1-148) x US-08-729-955A-2 (1-3258)

Qy         19 LeuGlyPheLysGluGlyGln---------IleGlnGlyThrLysAspGlnTyrLeuGlu 35
Db       2619 TTGAAATTTAAGAAATGGATGTGTGAGCTCATCCAGGGTCAAAAGCAAGGCTTCCAAGAA 2678

Qy         36 GlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGln 55
Db       2679 GATGCAGAGATGAACAAGAGAAATGAAGAAAAGAAATTTGGAATTTCAGGAAGTCACGAA 2738

Qy         56 GluLeuMetLysPhe---------------TrpLeuSerHisIleAspGlnTyrAsnAsn 70
Db       2739 GACCTTTCAAAATTTTCACTTGACAAAAATCAGTTGGCACACAACAAACAATCAAGTACA 2798

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIle 90
Db       2799 AGGAGCTCAGAAGATTATCATTTAAAT--------------------------AGTTTC 2831

Qy         91 ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys 110
Db       2832 AGTAACCCTCCAAGACAA---------TATCAGAAAATCATGAAGAGACTCATTAAAAGA 2882

Qy        111 LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu 130
Db       2883 TATGTATTGCAGGCCCAGATTGATAAGGAGAGCGATGAGGTGAAT------GAAGGGGAA 2936

Qy        131 ValLysGluValGlyGlyThrLeu--------------------------GlnValSer 141
Db       2937 TTGAAGGAAATTAAGCAAGACATCTCAAGTCTCCGTTATGAACTCCTTGAAGAGAAATCA 2996

Qy        142 GluAsnProAspAspMet 147
Db       2997 CAGAACTCAGAAGACCTA 3014

RESULT 5
US-09-134-001C-716
; Sequence 716, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 716
;   LENGTH: 2271
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-716

Alignment Scores:
Pred. No.:                0.828          Length:        2271
Score:                    82.00          Matches:       35
Percent Similarity:       43.42%         Conservative:  31
```

```
Best Local Similarity:  23.03%          Mismatches:     58
Query Match:            10.62%          Indels:         28
DB:                     4               Gaps:           6

US-09-980-054A-12 (1-148) x US-09-134-001C-716 (1-2271)

Qy          7 AspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIle 26
              ::::::|||:::   |||  ||||||     |||                    |||
Db        649 GAAGATGTGATTCGTTTACGTGAAGAATTATCAGAA-----------------CAATAT 690

Qy         27 GlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGln------------ 42
              :::     :::|||              |||::::::||||||:::
Db        691 CGCTCACTTCAAGATTTAAAAGAATTAGGACAAAAATATGGATTTGATATTAGCCGTCCT 750

Qy         43 ---ThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrp 61
                 |||   |||:::                        :::|||   |||         :::
Db        751 GCTACTAACTTCAAAGAA--------------GCTGTGCAATGGTTATACTTAGCATAT 795

Qy         62 LeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeu 81
              |||:::   |||   :::   |||   ::::::                    :::
Db        796 TTAGCTGCTATCAAAGAACAAAATGGTGCAGCAATGAGTTTAGGACGTACTTCAACATTC 855

Qy         82 GluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGlu 101
              ||||||   ||:::      :::   |||||||||          ::::::   :::
Db        856 TTAGATATTTATGCTGAACGTGATTTACAAAATGGTGACATCACTGAACAAGAAGTTCAA 915

Qy        102 LysAsnIleLysLysAlaArgAsnLysLeuArgValIleAla---SerIleThrLysGlu 120
              :::   |||                |||||||||::::::      :::   |||   |||
Db        916 GAAATCATTGACCACTTCATTATGAAATTACGTATCGTTAAATTCGCACGTACGCCTGAA 975

Qy        121 --------------------------ThrTrpLysIleAspSerLeuAspAsnLeuVal 131
                                        ||||||      :::|||:::      :::
Db        976 TATAATGAATTATTCTCTGGAGATCCAACTTGGGTAACTGAATCTATCGGTGGTGTA--- 1032

Qy        132 LysGluValGlyGlyThrLeuGlnValSerGluAsn 143
                  :::   |||         |||::::::|||
Db       1033 ---GGTATTGACGGCCGTCCGATGGTAACTAAAAAC 1065

RESULT 6
US-09-508-370A-11
; Sequence 11, Application US/09508370A
; Patent No. 6492131
; GENERAL INFORMATION:
;  APPLICANT:      Dieter Soll
;  APPLICANT:  Michael Ibba
;  TITLE OF INVENTION:      Class I-type Lysyl-tRNA Synthetase
;  FILE REFERENCE:      OCR-896
;  CURRENT APPLICATION NUMBER: US/09/508,370A
;  CURRENT FILING DATE:  2000-03-10
;  PRIOR APPLICATION NUMBER:      PCT/US98/18968
;  PRIOR FILING DATE:      1998-09-09
;  NUMBER OF SEQ ID NOS:      16
;  SOFTWARE:      MS Dos
; SEQ ID NO 11
;   LENGTH: 1563
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: lysyl t-RNA synthetase
;   OTHER INFORMATION: clone described in Example 3
US-09-508-370A-11

Alignment Scores:
Pred. No.:              0.663           Length:         1563
Score:                  81.00           Matches:        32
Percent Similarity:     36.24%          Conservative:   22
Best Local Similarity:  21.48%          Mismatches:     49
Query Match:            10.49%          Indels:         46
DB:                     4               Gaps:           4

US-09-980-054A-12 (1-148) x US-09-508-370A-11 (1-1563)

Qy         14 GluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyr 33
              :::|||:::   :::      ||||||      :::|||                  |||
Db       1033 AAAGAAGAAAAAAAAAAGAGCATTTAAA------AGAATTTACGAACTATCTCAACCATAC 1086

Qy         34 LeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyr 53
              :::      |||        ||||||  ||||||:::        :::|||
Db       1087 ATGCCAAGCAAAAGAATCCCTTATCAGGTCGGATTCAGACATTTAAGTGTAATCAGTCAA 1146

Qy         54 IleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSer 73
              |||   |||                                      ||||||   :::
Db       1147 ATATTTGAA-----------------------------------AATAATATAAATAAA 1170

Qy         74 LeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAla-------------------- 86
              ::: - |||:::|||   |||:::::::::
Db       1171 ATTTTAAATTACTTGAAAAACGTTCAAGAAGATCAAAAAGACAAACTAATAAATAAAATA 1230

Qy         87 ---------------------------------------------GlnIleSerIleThr 91
                                                          :::   |||:::
Db       1231 AATTGCGCAATTAATTGGATAAGAGATTTTGCACCCGAAGATTTCAAATTTTCATTAAGA 1290

Qy         92 AsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeu 111
              :::            :::|||      |||:::|||   |||||||||   |||:::|||
Db       1291 TCTAAATTTGATAATATGGAAATACTAGAAGAAAATAGCAAAAAAGCAATTAATGAACTT 1350

Qy        112 ----------------------------ArgValIleAlaSerIleThrLysGluThr 121
                                                                |||      |||
Db       1351 TTGGATTTTTTAAAGAAAAATTTTGAAGTTGCCACAGAACAAGACATTCAAAACGAAATA 1410

Qy        122 TrpLysIleAspSerLeuAspAsnLeu 130
              :::||||||          :::|||:::
Db       1411 TATAAAATTTCAAGAGAAAATAATATA 1437

RESULT 7
US-09-166-350-27
; Sequence 27, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew
;  APPLICANT: Chen, Yao
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Old, Lloyd
;  APPLICANT: Jager, Elke
;  APPLICANT: Knuth, Alex
;  TITLE OF INVENTION: Renal Cancer Associated Antigens and
;  TITLE OF INVENTION: Uses Therefor
;  FILE REFERENCE: L0461/7051
;  CURRENT APPLICATION NUMBER: US/09/166,350A
;  CURRENT FILING DATE: 1998-10-05
;  EARLIER APPLICATION NUMBER: US 09/166,350
;  EARLIER FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27
;   LENGTH: 6773
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-166-350-27

Alignment Scores:
Pred. No.:              6.71            Length:         6773
Score:                  80.00           Matches:        35
Percent Similarity:     40.44%          Conservative:   20
Best Local Similarity:  25.74%          Mismatches:     54
Query Match:            10.36%          Indels:         27
DB:                     4               Gaps:           4

US-09-980-054A-12 (1-148) x US-09-166-350-27 (1-6773)

Qy          9 ValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGly 28
              |||      |||||||||   |||   :::|||      ::::::   |||:::|||
Db       3311 GTAAATGAACTTGAAGAACTTCAGGTACAACTTCAAAAGGAAAAGAAACAGCTTCAGAAA 3370

Qy         29 ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPhe 48
              |||      :::            |||:::         ||||||
```

```
Db       3371 ACCATGCAAGAATTAGAGCTGGTTAAAAAGGATGCCCAACAAACCACATTGATGAATATG 3430

Qy         49 LeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyr 68

Db       3431 GAAATAGCTGATTAT---GAACGTTTGATGAAA-------------------------- 3460

Qy         69 AsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle 88

Db       3461 GAACTAAATCAAAAGTTAACTAATAAAAACAACAAGATAGAAGATTTGGAGCAAGAAATA 3520

Qy         89 ----------------------------------SerIleThrAsnGlyAspLysGlu 96

Db       3521 AAAATTCAAAAACAGAAACAAGAAACCCTACAAGAAGAAATAACTTCATTACAGTCTTCA 3580

Qy         97 ValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSer 116

Db       3581 GTACAACAATATGAAGAAAAA----------AACACCAAAATCAAGCAATTGCTTGTG 3628

Qy        117 IleThrLysGlu-ThrTrpLysIleAspSerLeuAspAsnLeuVal 131

Db       3629 AAAACCAAAAAGGAACTGGCAGATTCAAAGCAAGCAGAAACTGATC 3674
```

```
RESULT 8
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
;   LENGTH: 1230025
;   TYPE: DNA
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(15000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (15001)..(30000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (30001)..(45000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (45001)..(60000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (60001)..(75000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (75001)..(90000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (90001)..(105000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (105001)..(120000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (120001)..(135000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (135001)..(150000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (150001)..(165000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (165001)..(180000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (180001)..(195000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (195001)..(210000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (210001)..(225000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (225001)..(240000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (240001)..(255000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (255001)..(270000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (270001)..(285000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (285001)..(300000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (300001)..(315000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (315001)..(330000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (330001)..(345000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (345001)..(360000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (360001)..(375000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (375001)..(390000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (390001)..(405000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (405001)..(420000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (420001)..(435000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (435001)..(450000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (450001)..(465000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (465001)..(480000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (480001)..(495000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (495001)..(510000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (510001)..(525000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
```

; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Alignment Scores:

Pred. No.:	8.68e+03	Length:	1230025
Score:	80.00	Matches:	39
Percent Similarity:	39.86%	Conservative:	16
Best Local Similarity:	28.26%	Mismatches:	47
Query Match:	10.36%	Indels:	36
DB:	4	Gaps:	6

US-09-980-054A-12 (1-148) x US-09-198-452A-1 (1-1230025)

```
Qy          7 AspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIle 26
              |||   ||||||   |||:::                          |||   |||:::
Db     199813 GACGAAGTCCTTGCATTGAAGAGTACAATG-----------------GAAGCTCAGCTT 199854

Qy         27 GlnGlyThrLysAspGlnTyr------LeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
              ||||||   ||||||               |||||||||
Db     199855 CAAGGATTTAAGGACCTTGTAGGTACCTGGGAAGGAAAA--------------------- 199893

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHis 64
                                           ||||||       |||       ||||||
Db     199894 -------------------------TATCAGGAATTTAAGAAAAACAAGCTTTCCAAA 199926

Qy         65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
              :::       |||:::   :::   |||       |||   ||||||   ||||||   :::
Db     199927 GTTTTAGTTTACGACTTCACAAAATCCTATTCTAACCTTCTAAATCGTTTGGAGGTACTC 199986

Qy         85 MetAlaGlnIleSer-----------IleThrAsnGlyAspLysGluValGluAspTyr 100
                 |||:::   |||           :::   :::   |||:::       ||||||
Db     199987 CATGCCGAGAGCTCCACGGATGATTTGGTATTACATGTCGATAGAATGTCGGAAGATCTG 200046

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              :::|||   |||::::::           |||                ::::||   |||
Db     200047 AAGAAAACAATCGAGGAGATTGACGGCAATTTA---------TTTCAGGTAACTCCTGAA 200097

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeu 138
                                 |||       |||   :::|||       |||   :::
Db     200098 --------------GAGCTCTCTTTGTTAGCTCGGGAATATCAGGGACTCATG 200136
```

RESULT 9
US-09-252-991A-6512
; Sequence 6512, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6512
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6512

Alignment Scores:

Pred. No.:	0.259	Length:	417
Score:	78.00	Matches:	29
Percent Similarity:	38.93%	Conservative:	22
Best Local Similarity:	22.14%	Mismatches:	34
Query Match:	10.10%	Indels:	46
DB:	4	Gaps:	5

```
US-09-980-054A-12 (1-148) x US-09-252-991A-6512 (1-417)

Qy          29 ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPhe 48
               :::::::::|||   :::|||        :::|||        |||         |||
Db          78 TCCCGCGAGCAGGTGGTCGAGCAACTGCACTTCGGT------GGCGGAACCCCGACCTTC 131

Qy          49 LeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyr 68
               |||     |||   ::::::|||||||||
Db         132 CTCAGTCCCGGCCAATTGCGCGAACTCATG---------------------------- 161

Qy          69 AsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp-------------- 83
                   |||   ||||||   |||||||||   |||:::|||
Db         162 ---------TCGCAATTGCGCACCCACCTCAACCTGCTGGACGACGATTCGGGCGACTAC 212

Qy          84 ---------------------------------------------IleMetAlaGlnIle 88
                                                           ::::::   ::::::
Db         213 GGCATCGAGATCGACCCGCGCGAGGCCGACTGGTCGACCATGGGCCTGCTCCGCGAACTC 272

Qy          89 SerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108
                           |||:::|||::::::      ::::::||||||
Db         273 GGTTTCAACCGCGTCAGCCTCGGCGTACAGGACTTCGACATGGAGGTGCAGAAGGCGGTC 332

Qy         109 AsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAsp 128
               |||:::::::::                :::||||||
Db         333 AACCGCATGCAGACC------------CCGGAGGAGACC-------------------- 359

Qy         129 AsnLeuValLysGluValGlyGlyThrLeuGln 139
                     :::   |||         |||||||||
Db         360 CGTACCATCGTCGAGGCCGCGCGCACCCTGCAA 392

RESULT 10
US-09-252-991A-6539/c
; Sequence 6539, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6539
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6539

Alignment Scores:
Pred. No.:              0.579          Length:         747
Score:                  78.00          Matches:        29
Percent Similarity:     38.93%         Conservative:   22
Best Local Similarity:  22.14%         Mismatches:     34
Query Match:            10.10%         Indels:         46
DB:                     4              Gaps:           5

US-09-980-054A-12 (1-148) x US-09-252-991A-6539 (1-747)

Qy          29 ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPhe 48
               :::::::::|||   :::|||        :::|||        |||         |||
Db         402 TCCCGCGAGCAGGTGGTCGAGCAACTGCACTTCGGT------GGCGGAACCCCGACCTTC 349

Qy          49 LeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyr 68
               |||     |||   ::::::|||||||||
Db         348 CTCAGTCCCGGCCAATTGCGCGAACTCATG---------------------------- 319

Qy          69 AsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp-------------- 83
                   |||   ||||||   |||||||||   |||:::|||
Db         318 ---------TCGCAATTGCGCACCCACCTCAACCTGCTGGACGACGATTCGGGCGACTAC 268

Qy          84 ---------------------------------------------IleMetAlaGlnIle 88
                                                           ::::::   ::::::
Db         267 GGCATCGAGATCGACCCGCGCGAGGCCGACTGGTCGACCATGGGCCTGCTCCGCGAACTC 208

Qy          89 SerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108
                           |||:::|||::::::      ::::::||||||
Db         207 GGTTTCAACCGCGTCAGCCTCGGCGTACAGGACTTCGACATGGAGGTGCAGAAGGCGGTC 148

Qy         109 AsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAsp 128
               |||:::::::::                :::||||||
Db         147 AACCGCATGCAGACC------------CCGGAGGAGACC-------------------- 121

Qy         129 AsnLeuValLysGluValGlyGlyThrLeuGln 139
                     :::   |||         |||||||||
Db         120 CGTACCATCGTCGAGGCCGCGCGCACCCTGCAA 88

RESULT 11
US-09-252-991A-6495
; Sequence 6495, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6495
;   LENGTH: 1521
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6495

Alignment Scores:
Pred. No.:              1.54           Length:         1521
Score:                  78.00          Matches:        29
Percent Similarity:     38.93%         Conservative:   22
Best Local Similarity:  22.14%         Mismatches:     34
Query Match:            10.10%         Indels:         46
DB:                     4              Gaps:           5

US-09-980-054A-12 (1-148) x US-09-252-991A-6495 (1-1521)

Qy          29 ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPhe 48
               :::::::::|||   :::|||        :::|||        |||         |||
Db         442 TCCCGCGAGCAGGTGGTCGAGCAACTGCACTTCGGT------GGCGGAACCCCGACCTTC 495

Qy          49 LeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyr 68
               |||     |||   ::::::|||||||||
Db         496 CTCAGTCCCGGCCAATTGCGCGAACTCATG---------------------------- 525

Qy          69 AsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp-------------- 83
                   |||   ||||||   |||||||||   |||:::|||
Db         526 ---------TCGCAATTGCGCACCCACCTCAACCTGCTGGACGACGATTCGGGCGACTAC 576

Qy          84 ---------------------------------------------IleMetAlaGlnIle 88
                                                           ::::::   ::::::
Db         577 GGCATCGAGATCGACCCGCGCGAGGCCGACTGGTCGACCATGGGCCTGCTCCGCGAACTC 636

Qy          89 SerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108
                           |||:::|||::::::      ::::::||||||
Db         637 GGTTTCAACCGCGTCAGCCTCGGCGTACAGGACTTCGACATGGAGGTGCAGAAGGCGGTC 696

Qy         109 AsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAsp 128
               |||:::::::::                :::||||||
Db         697 AACCGCATGCAGACC------------CCGGAGGAGACC-------------------- 723
```

```
Qy        129 AsnLeuValLysGluValGlyGlyThrLeuGln 139
                  :::  |||        |||||||||
Db        724 CGTACCATCGTCGAGGCCGCGCGCACCCTGCAA 756

RESULT 12
US-08-923-992A-3
; Sequence 3, Application US/08923992A
; Patent No. 6280738
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3312
US-08-923-992A-3

Alignment Scores:
Pred. No.:                4.5          Length:          3312
Score:                    78.00        Matches:         36
Percent Similarity:       41.33%       Conservative:    26
Best Local Similarity:    24.00%       Mismatches:      64
Query Match:              10.10%       Indels:          24
DB:                       3            Gaps:            5

US-09-980-054A-12 (1-148) x US-08-923-992A-3 (1-3312)

Qy          5 AspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGly 24
              |||:::||||||   :::        |||::::::  |||          |||
Db        268 GATGTGGATAACACAATTCTATCTCATGAACAGAAAAATGAGTTTAAAACAAAAATTGAT 327

Qy         25 GlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
                :::        :::   |||      ||||||   :::          :::|||
Db        328 GAAACAAATGATTCT---GATGCATTATTAGAATTAGAAAATCAATTTAACGAAACT--- 381

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeu---------------------- 57
                  |||   |||   |||         :::|||:::

Db        382 ---AATAGACTGTTACACATCAAACAACATGAAGAAGTTGAGAAACATAACAAACCTAAC 438

Qy         58 ------------MetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSer 73
                         :::|||         :::   :::|||      |||
Db        439 CAACAGAAAACTCTGAAACAGTCAGATACGAAAGTAGATCTAAGCAATATTGACAAAGAG 498

Qy         74 LeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGly 93
              |||   :::      :::   :::|||   :::      |||      |||||||||
Db        499 CTTAATCATCAAAAAAGTCAAGTTGAAGCAATGGCAGAGCAAGCGGGAATCACAAATGAA 558

Qy         94 AspLysGlu---------------ValGluAspTyrGluLysAsnIleLysLysAlaArg 108
              |||||:::                :::||||||      |||      :::::|||
Db        559 GATAAAGATTCTATGCTGAAAAAAATCGAAGATATTCGTAAACAAGCTCAACAAGCAGAT 618

Qy        109 AsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSer------ 126
                |||                  :::   :::|||      |||:::   |||
Db        619 AAAAAAGAAGATGCCGAAGTAAAGGTTCGTGAAGAACTAGGTAAACTCTTTAGTTCAACT 678

Qy        127 ---------LeuAspAsnLeuValLysGlu 133
                       ||||||     ::::::|||
Db        679 AAAGCTGGTCTGGATCAACAAATTCAAGAG 708

RESULT 13
US-08-867-611-9
; Sequence 9, Application US/08867611
; Patent No. 6172189
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, SUSHIL G
;    APPLICANT:  DESAI, SURESH M
;    APPLICANT:  CASEY, JAMES M
;    APPLICANT:  DAILEY, STEPHEN H
;    APPLICANT:  DAWSON, GEORGE J
;    APPLICANT:  GUTIERREZ, ROBIN A
;    APPLICANT:  LESNIEWSKI, RICHARD R
;    APPLICANT:  STEWART, JAMES L
;    APPLICANT:  RUPPRECHT, KEVIN R
;    TITLE OF INVENTION:  HEPATITIS C ASSAY UTILIZING RECOMBINANT
;    TITLE OF INVENTION:  ANTIGENS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD, CHAD377/AP6D2
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,611
;      FILING DATE:  02-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,757
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/179,896
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/572,822
;      FILING DATE:  24-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/614,069
;      FILING DATE:  07-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,561
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,565
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 07/748,566
;      FILING DATE:  21-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834.US.P6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1488 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1488
US-08-867-611-9

Alignment Scores:
Pred. No.:              1.73          Length:        1488
Score:                  77.50         Matches:       37
Percent Similarity:     32.60%        Conservative:  22
Best Local Similarity:  20.44%        Mismatches:    67
Query Match:            10.04%        Indels:        55
DB:                     3             Gaps:          6

US-09-980-054A-12 (1-148) x US-08-867-611-9 (1-1488)

Qy         12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
              |||:::   |||   |||   |||   |||             |||          :::|||
Db        436 CTCGACGCTGAAGGGTATGCACTGTACTTCTCTCGCGCCACCATTCCTTGGGATCGTGAT 495

Qy         32 GlnTyrLeuGluGlyLysGlu----------------------------------Tyr 39
              ::::::   ||||||   |||                                   |||
Db        496 CGTTTTGCAGAAGGCCTTGAAACCGTTGGCGATAACTTCCTGCGTCATCTTGGTATTTAT 555

Qy         40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle-------------------------- 50
              ||||||:::   ||||||   |||      :::
Db        556 GGCTACCGTGCAGGCTTTATCCGTCGTTACGTCAACTGGCAGCCAAGTCCGTTAGAACAC 615

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer---------HisIle------ 65
              |||       :::::::::|||         |||                |||:::
Db        616 ATCGAAATGTTAGAGCAGCTTCGTGTTCTGTGGTACGGCGAAAAAATCCATGTTGCTGTT 675

Qy         66 -------------------------------------AspGlnTyrAsnAsnSerSer 72
                                                   |||          :::::::::
Db        676 GCTCAGGAAGTTCCTGGCACAGGTGTGGATACCCCTGAAGATCTCGACCCGTCGACGAAT 735

Qy         73 SerLeuArgAsnHisLeuAsn---------------------------------------- 79
              ||||||   :::|||
Db        736 TCTCTAGACTCCCACTACCAGGACGTTCTGAAAGAAGTTAAAGCTGCTGCTTCTAAAGTT 795

Qy         80 ------AsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluVal 97
                    ||||||     :::          |||:::|||
Db        796 AAAGCTAACCTGCTGTCTGTTGAAGAAGCATGCTCTCTGACCCCGCCGCACTCTGCTAAA 855

Qy         98 GluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
                 :::       |||     |||       |||           :::
Db        856 TCTAAATTCGGTTACGGTGCTAAAGACGTTCGTTGCCACGCTCGTAAAGCTGTTACCCAC 915

Qy        118 ThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThr 137
                  ||||||   |||   |||:::::   |||       :::       |||
Db        916 ATCAACTCTGTTTGGAAA---GATCTGCTGGAAGACAACGTTACCCCGATCGACACCACC 972

Qy        138 Leu 138
              :::
Db        973 ATC 975

RESULT 14
PCT-US92-06965A-14
```

```
; Sequence 14, Application PC/TUS9206965A
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, S.
;    APPLICANT:  DESAI, S.
;    APPLICANT:  DAILEY, S.
;    TITLE OF INVENTION:  HCV SYNTHETIC PEPTIDE FROM NS1 REGION
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  U.S.
;      ZIP:  60065-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06965A
;      FILING DATE:  19920821
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834PC.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1488 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1488
PCT-US92-06965A-14

Alignment Scores:
Pred. No.:              1.73          Length:        1488
Score:                  77.50         Matches:       37
Percent Similarity:     32.60%        Conservative:  22
Best Local Similarity:  20.44%        Mismatches:    67
Query Match:            10.04%        Indels:        55
DB:                     5             Gaps:          6

US-09-980-054A-12 (1-148) x PCT-US92-06965A-14 (1-1488)

Qy         12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
              |||:::   |||   |||   |||   |||             |||          :::|||
Db        436 CTCGACGCTGAAGGGTATGCACTGTACTTCTCTCGCGCCACCATTCCTTGGGATCGTGAT 495

Qy         32 GlnTyrLeuGluGlyLysGlu----------------------------------Tyr 39
              ::::::   ||||||   |||                                   |||
Db        496 CGTTTTGCAGAAGGCCTTGAAACCGTTGGCGATAACTTCCTGCGTCATCTTGGTATTTAT 555

Qy         40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle-------------------------- 50
              ||||||:::   ||||||   |||      :::
Db        556 GGCTACCGTGCAGGCTTTATCCGTCGTTACGTCAACTGGCAGCCAAGTCCGTTAGAACAC 615

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer---------HisIle------ 65
              |||       :::::::::|||         |||                |||:::
Db        616 ATCGAAATGTTAGAGCAGCTTCGTGTTCTGTGGTACGGCGAAAAAATCCATGTTGCTGTT 675

Qy         66 -------------------------------------AspGlnTyrAsnAsnSerSer 72
                                                   |||          :::::::::
Db        676 GCTCAGGAAGTTCCTGGCACAGGTGTGGATACCCCTGAAGATCTCGACCCGTCGACGAAT 735

Qy         73 SerLeuArgAsnHisLeuAsn---------------------------------------- 79
```

```
                ||||||   :::|||
Db        736 TCTCTAGACTCCCACTACCAGGACGTTCTGAAAGAAGTTAAAGCTGCTGCTTCTAAAGTT 795

Qy         80 ------AsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluVal 97
                  ||||||      :::          |||:::|||
Db        796 AAAGCTAACCTGCTGTCTGTTGAAGAAGCATGCTCTCTGACCCCGCCGCACTCTGCTAAA 855

Qy         98 GluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
                 :::          |||     |||      |||         :::
Db        856 TCTAAATTCGGTTACGGTGCTAAAGACGTTCGTTGCCACGCTCGTAAAGCTGTTACCCAC 915

Qy        118 ThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThr 137
                      ||||||   |||   |||::::::   |||       :::       |||
Db        916 ATCAACTCTGTTTGGAAA---GATCTGCTGGAAGACAACGTTACCCCGATCGACACCACC 972

Qy        138 Leu 138
                :::
Db        973 ATC 975
```

```
RESULT 15
US-08-500-222-1
; Sequence 1, Application US/08500222
; Patent No. 5643733
;  GENERAL INFORMATION:
;    APPLICANT:  Robinson, John M
;    APPLICANT:  Pilot-Matias, Tami J
;    APPLICANT:  Hunt, Jeffrey C
;    TITLE OF INVENTION:  Borrelia burgdorferi antigens
;    TITLE OF INVENTION:  and uses thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  One Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/500,222
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/779,704
;      FILING DATE:  21-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wong, Wean Khing
;      REGISTRATION NUMBER:  33561
;      REFERENCE/DOCKET NUMBER:  5051.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-937-9556
;   INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1497 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Borrelia burgdorferi
;      STRAIN:  B31
;    IMMEDIATE SOURCE:
;      CLONE:  pB776
US-08-500-222-1

Alignment Scores:
```

```
Pred. No.:                1.75          Length:          1497
Score:                    77.50         Matches:         34
Percent Similarity:       35.47%        Conservative:    27
Best Local Similarity:    19.77%        Mismatches:      46
Query Match:              10.04%        Indels:          65
DB:                       1             Gaps:            8

US-09-980-054A-12 (1-148) x US-08-500-222-1 (1-1497)

Qy         12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
              |||:::   |||   |||   |||   |||           |||         :::|||
Db        436 CTCGACGCTGAAGGGTATGCACTGTACTTCTCTCGCGCCACCATTCCTTGGGATCGTGAT 495

Qy         32 GlnTyrLeuGluGlyLysGlu----------------------------------Tyr 39
              ::::::   ||||||  |||                                    |||
Db        496 CGTTTTGCAGAAGGCCTTGAAACCGTTGGCGATAACTTCCTGCGTCATCTTGGTATTTAT 555

Qy         40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle-------------------------- 50
              ||||||:::   ||||||   |||     :::
Db        556 GGCTACCGTGCAGGCTTTATCCGTCGTTACGTCAACTGGCAGCCAAGTCCGTTAGAACAC 615

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer---------HisIle------ 65
              |||      :::::::::|||        |||              |||:::
Db        616 ATCGAAATGTTAGAGCAGCTTCGTGTTCTGTGGTACGGCGAAAAAATCCATGTTGCTGTT 675

Qy         65 ------------------------------------------------------------ 65

Db        676 GCTCAGGAAGTTCCTGGCACAGGTGTGGATACCCCTGAAAATCCGTCGACAGGGCTTATG 735

Qy         66 ---------AspGlnTyrAsnAsnSerSerSerLeu------------------ArgAsn 76
                      |||      |||   |||   ::::::
Db        736 AAGATCTCAGACCCGAGAAATACTTCAAAGGCTATTAATTTTATTCAGACAACAGAAGGG 795

Qy         77 HisLeuAsnAsnLeuGluAspIle---------MetAlaGlnIleSerIleThrAsnGly 93
              :::||||||   :::|||   :::         |||   ::::::::::::   :::|||
Db        796 AATTTAAATGAAGTAGAAAAAGTCTTAGTAAGAATGAAGGAATTGGCAGTTCAATCAGGT 855

Qy         94 AspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgVal 113
              :::          |||   :::                  |||       :::::::::
Db        856 AACGGCACATATTCAGATGCAGAC-----------------AGAGGTTCTATACAAATT 897

Qy        114 ---IleAlaSerIleThrLysGluThrTrpLysIle 124
                 |||      :::|||   |||      :::|||
Db        898 GAAATAGAGCAACTTACAGACGAAATTAATAGAATT 933
```

```
Search completed: February 16, 2004, 16:23:50
Job time : 955 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 15:27:53 ; Search time 332 Seconds
(without alignments)
1642.101 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980054/runat_13022004_163843_26438/app_query.fasta_1.327
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980054_@CGN_1_1_391_@runat_13022004_163843_26438
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	764	99.0	447	13	US-10-032-585-6062	Sequence 6062, Ap
c 2	119.5	15.5	2877	15	US-10-128-714-6171	Sequence 6171, Ap
c 3	119.5	15.5	3863	15	US-10-128-714-171	Sequence 171, App
c 4	119.5	15.5	4877	15	US-10-128-714-5171	Sequence 5171, Ap
c 5	107	13.9	65	13	US-10-032-585-2884	Sequence 2884, Ap
c 6	106.5	13.8	2601	15	US-10-128-714-7171	Sequence 7171, Ap
7	99.5	12.9	490	12	US-10-242-535A-56991	Sequence 56991, A
8	91.5	11.9	3018	12	US-10-369-493-24748	Sequence 24748, A
c 9	91	11.8	55	13	US-10-032-585-3062	Sequence 3062, Ap
10	87	11.3	5967	15	US-10-171-581-158	Sequence 158, App
11	86	11.1	7100	10	US-09-932-183A-1	Sequence 1, Appli
c 12	85.5	11.1	6723	10	US-09-070-927A-163	Sequence 163, App
13	85	11.0	255	12	US-10-369-493-44827	Sequence 44827, A
14	83.5	10.8	2637	12	US-10-369-493-44024	Sequence 44024, A
15	83.5	10.8	2640	12	US-10-369-493-24958	Sequence 24958, A
16	83.5	10.8	2643	12	US-10-369-493-45330	Sequence 45330, A
17	83.5	10.8	2816	13	US-10-094-749-289	Sequence 289, App
18	83.5	10.8	3261	15	US-10-005-216-1	Sequence 1, Appli
19	83	10.8	3552	9	US-09-815-242-4174	Sequence 4174, Ap
20	83	10.8	3567	9	US-09-815-242-8027	Sequence 8027, Ap
21	82.5	10.7	1308	13	US-10-032-585-6260	Sequence 6260, Ap
22	82.5	10.7	3678	12	US-10-369-493-45562	Sequence 45562, A
c 23	82	10.6	507	10	US-09-796-692-4355	Sequence 4355, Ap
c 24	82	10.6	507	12	US-10-057-475B-4355	Sequence 4355, Ap
c 25	82	10.6	507	12	US-10-154-884B-4355	Sequence 4355, Ap
c 26	82	10.6	507	15	US-10-040-862-4355	Sequence 4355, Ap
27	82	10.6	510	11	US-09-918-995-24731	Sequence 24731, A
28	81	10.5	963	9	US-09-815-242-9687	Sequence 9687, Ap
29	81	10.5	3491	8	US-08-781-986A-117	Sequence 117, App
30	80.5	10.4	765	13	US-10-032-585-6511	Sequence 6511, Ap
31	80	10.4	3471	12	US-10-369-493-23730	Sequence 23730, A
32	80	10.4	6773	10	US-09-864-864-336	Sequence 336, App
33	80	10.4	1230025	12	US-10-289-762-1	Sequence 1, Appli
34	79.5	10.3	1319	13	US-09-814-353-21904	Sequence 21904, A
35	79.5	10.3	1934	13	US-09-882-227-201	Sequence 201, App
c 36	79.5	10.3	684707	12	US-10-398-221-9	Sequence 9, Appli
c 37	79.5	10.3	3011208	12	US-10-398-221-2058	Sequence 2058, Ap
38	79	10.2	3522	12	US-10-369-493-45641	Sequence 45641, A
39	78.5	10.2	1420	12	US-10-014-099F-62	Sequence 62, Appl
40	78.5	10.2	1441	12	US-10-014-099F-66	Sequence 66, Appl
41	78.5	10.2	2712	12	US-10-369-493-24735	Sequence 24735, A
42	78.5	10.2	5290	12	US-10-014-099F-76	Sequence 76, Appl
43	78.5	10.2	5309	12	US-10-014-099F-77	Sequence 77, Appl
44	78	10.1	1125	13	US-10-032-585-6585	Sequence 6585, Ap
45	77.5	10.0	369	12	US-10-242-535A-42952	Sequence 42952, A

ALIGNMENTS

RESULT 1
US-10-032-585-6062
; Sequence 6062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6062
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6062

Alignment Scores:
Pred. No.: 9.01e-87 Length: 447
Score: 764.00 Matches: 146

```
Percent Similarity:     100.00%         Conservative:  2
Best Local Similarity:  98.65%          Mismatches:    0
Query Match:            98.96%          Indels:        0
DB:                     13              Gaps:          0

US-09-980-054A-12 (1-148) x US-10-032-585-6062 (1-447)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              ||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        141 SerGluAsnProAspAspMetTrp 148
              |||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGG 444

RESULT 2
US-10-128-714-6171/c
; Sequence 6171, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6171
;   LENGTH: 2877
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-6171

Alignment Scores:
Pred. No.:              0.00017         Length:        2877
Score:                  119.50          Matches:       34
Percent Similarity:     40.67%          Conservative:  27
Best Local Similarity:  22.67%          Mismatches:    48
Query Match:            15.48%          Indels:        41
DB:                     15              Gaps:          3

US-09-980-054A-12 (1-148) x US-10-128-714-6171 (1-2877)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::     |||
Db        539 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 480

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::      |||   ||||||      :::      :::   ||||||
Db        479 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 420

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||   :::|||   :::                 |||       :::
Db        419 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 360

Qy         65 ------------------------------------------------------------ 65

Db        359 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 300

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                    ||||||
Db        299 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 240

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                      :::|||||||   |||      :::|||      |||   ::::::::
Db        239 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 180

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                 |||   :::   :::   |||   ::::::   |||   |||   ::::::   :::
Db        179 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 120

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
              |||   |||       |||   ::::::|||
Db        119 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 90

RESULT 3
US-10-128-714-171/c
; Sequence 171, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 171
;   LENGTH: 3863
;   TYPE: DNA
```

```
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-171

Alignment Scores:
Pred. No.:              0.000261        Length:         3863
Score:                  119.50          Matches:        34
Percent Similarity:     40.67%          Conservative:   27
Best Local Similarity:  22.67%          Mismatches:     48
Query Match:            15.48%          Indels:         41
DB:                     15              Gaps:           3

US-09-980-054A-12 (1-148) x US-10-128-714-171 (1-3863)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::      |||
Db        677 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 618

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::      |||   ||||||      :::      :::   ||||||
Db        617 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 558

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||   :::|||   :::                  |||         :::
Db        557 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 498

Qy         65 ------------------------------------------------------------ 65

Db        497 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 438

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                 ||||||
Db        437 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 378

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                    :::||||||   |||      :::|||      |||   :::::::::
Db        377 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 318

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                |||   :::   :::   |||   ::::::   |||   |||   ::::::   :::
Db        317 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 258

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
                |||   |||       |||   :::::|||
Db        257 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 228

RESULT 4
US-10-128-714-5171/c
; Sequence 5171, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5171
;   LENGTH: 4877
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-5171

Alignment Scores:
Pred. No.:              0.000367        Length:         4877
Score:                  119.50          Matches:        34
Percent Similarity:     40.67%          Conservative:   27
Best Local Similarity:  22.67%          Mismatches:     48
Query Match:            15.48%          Indels:         41
DB:                     15              Gaps:           3

US-09-980-054A-12 (1-148) x US-10-128-714-5171 (1-4877)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::      |||
Db       1539 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 1480

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::      |||   ||||||      :::      :::   ||||||
Db       1479 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 1420

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||   :::|||   :::                  |||         :::
Db       1419 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 1360

Qy         65 ------------------------------------------------------------ 65

Db       1359 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 1300

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                 ||||||
Db       1299 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 1240

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                    :::||||||   |||      :::|||      |||   :::::::::
Db       1239 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 1180

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                |||   :::   :::   |||   ::::::   |||   |||   ::::::   :::
Db       1179 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 1120

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
                |||   |||       |||   :::::|||
Db       1119 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 1090

RESULT 5
US-10-032-585-2884/c
; Sequence 2884, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2884
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-2884

Alignment Scores:
Pred. No.:              2.59e-05        Length:         65
Score:                  107.00          Matches:        21
Percent Similarity:     100.00%         Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            13.86%         Indels:        0
DB:                     13             Gaps:          0

US-09-980-054A-12 (1-148) x US-10-032-585-2884 (1-65)

Qy          2 SerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPhe 21
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 6

Qy         22 Lys 22
              |||
Db          5 AAA 3

RESULT 6
US-10-128-714-7171/c
; Sequence 7171, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Hu, Wenqi
;  APPLICANT:  Tishkoff, Daniel
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Eroshkin, Alexey M
;  APPLICANT:  Lemieux, Sebastien M
;  TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: 10182-018-999
;  CURRENT APPLICATION NUMBER: US/10/128,714
;  CURRENT FILING DATE:  2002-04-23
;  PRIOR APPLICATION NUMBER: US 60/285,697
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/287,066
;  PRIOR FILING DATE: 2001-04-27
;  PRIOR APPLICATION NUMBER: US 60/295,890
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/303,899
;  PRIOR FILING DATE: 2001-07-09
;  PRIOR APPLICATION NUMBER: US 60/316,362
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 8603
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7171
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: Aspergillus fumigatus
US-10-128-714-7171

Alignment Scores:
Pred. No.:              0.00636        Length:        2601
Score:                  106.50         Matches:       35
Percent Similarity:     42.66%         Conservative:  26
Best Local Similarity:  24.48%         Mismatches:    48
Query Match:            13.80%         Indels:        35
DB:                     15             Gaps:          3

US-09-980-054A-12 (1-148) x US-10-128-714-7171 (1-2601)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::     |||
Db        517 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 458

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::     |||   |||||     :::     :::   ||||||
Db        457 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 398

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeu--------- 62
              ::::::|||||   :::|||   :::             |||
Db        397 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 338

Qy         63 ------SerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHis--------- 77
                    |||   |||   :::   ::::::     |||   ||||||
Db        337 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGA-CGGAACAGATACCCGCAA 279

Qy         77 ---------------------------------------------------------- 77

Db        278 ATCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCG 219

Qy         78 ---LeuAsnAsnLeuGluAspIleMet--------------AlaGlnIleSerIleThr 91
                 |||||||||   ::::::                  |||   :::   :::
Db        218 CTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGATCCAGCTTCGTTGCCTATGGAG 159

Qy         92 AsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeu 111
              |||   ::::::   |||   |||   ::::::   :::|||   |||       |||
Db        158 AACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTTAAGGGCGCTGCAATCAAGGCC 99

Qy        112 ArgValIle 114
                 ::::::|||
Db         98 AAGCTTATC 90

RESULT 7
US-10-242-535A-56991
; Sequence 56991, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
;  APPLICANT: ChondroGene Inc.
;  APPLICANT:  Liew, C.C.
;  TITLE OF INVENTION: Compositions and Methods Relatiing to Osteoarthritis
;  FILE REFERENCE: 4231/2005
;  CURRENT APPLICATION NUMBER: US/10/242,535A
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 10/085,783
;  PRIOR FILING DATE: 2002-02-28
;  PRIOR APPLICATION NUMBER: US 60/305,340
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: US 60/275,017
;  PRIOR FILING DATE: 2001-03-12
;  PRIOR APPLICATION NUMBER: US 60/271,955
;  PRIOR FILING DATE: 2001-02-28
;  NUMBER OF SEQ ID NOS: 58994
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56991
;   LENGTH: 490
;   TYPE: DNA
;   ORGANISM: Human
US-10-242-535A-56991

Alignment Scores:
Pred. No.:              0.00428        Length:        490
Score:                  99.50          Matches:       32
Percent Similarity:     45.83%         Conservative:  23
Best Local Similarity:  26.67%         Mismatches:    46
Query Match:            12.89%         Indels:        19
DB:                     12             Gaps:          3

US-09-980-054A-12 (1-148) x US-10-242-535A-56991 (1-490)

Qy         11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
              |||::::::           |||   |||::::::::::|||       |||
Db        139 AACATGCAAAGACGAGTCAAAGAA---GGTTATAGAGATGGAATAGATGCTGGCAAAGCA 195

Qy         31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 50
                    :::|||           ||||||:::   |||   :::       |||
Db        196 GTTACTCTTCAACAGGGCTTCAATCAAGGTTATAAGAAAGGTGCAGAAGTCATTTTAAAC 255

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
              |||   ::::::      :::         ||||||               :::|||||||
Db        256 TATGGACGACTCCGAGGAACATTGAGTGCTTTGCTCTCCTGGTGTCACCTTCATAATAAT 315

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIle 90
              :::|||:::|||   |||   :::|||||||||   |||   :::   |||
Db        316 AATTCAACTTTGATCAATAAAATAAACAATCTTCTGGATGCAGTTGGCCAG--------- 366

Qy         91 ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys 110
                                  |||:::|||
Db        367 -----------------TGTGAAGAGTAT-------------------------GTG 381
```

```
Qy        111 LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu 130
              |||:::   :::  |||||||||      :::      :::|||   ||||||::::::
Db        382 CTCAAACATCTGAAATCAATCACTCCACCGTCCCATGTTGTAGATTTATTGGACTCCATT 441
```

RESULT 8
US-10-369-493-24748
; Sequence 24748, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24748
; LENGTH: 3018
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
US-10-369-493-24748

Alignment Scores:

Pred. No.:	0.609	Length:	3018
Score:	91.50	Matches:	43
Percent Similarity:	44.25%	Conservative:	34
Best Local Similarity:	24.71%	Mismatches:	62
Query Match:	11.85%	Indels:	35
DB:	12	Gaps:	7

US-09-980-054A-12 (1-148) x US-10-369-493-24748 (1-3018)

```
Qy          3 AspIleAsp---IleAspAsnValLeuAsn---------------------------- 11
              |||||||||   :::|||   :::||||||
Db       1024 GATATTGATAATTTAGATACTCTATTAAATAAAATAAAAGATGAAATTGAAAGAGTGGAA 1083

Qy         12 ------------LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGln 27
                        |||  ||||||     |||      :::      :::|||:::
Db       1084 ACTATTAAAGATTTGCTTGAAGAACTTAAAAATCTAAATGAAGAGATTGAAAAAATTGAA 1143

Qy         28 GlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArg 47
                 |||          |||   |||||||||                 :::::::::
Db       1144 AAATATAAAAGAATATGTGAAGAGTGCAAAGAATAC--------------TATGAGAAG 1188

Qy         48 PheLeu---IleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAsp 66
              :::|||   :::              |||      |||      |||   :::|||
Db       1189 TATTTAGAATTAGAAGAAAAGGCTGTAGAATACAATAAACTAACTTTAGAGTATATAACA 1248

Qy         67 GlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAla 86
                           |||:::     ::::::|||:::||||||      :::
Db       1249 TTGCTTCAGGAGAAAAAATCCATTGAAAAAAATATTAACGATTTAGAAACAAGAATTAAT 1308

Qy         87 GlnIle----------------------------SerIleThrAsnGlyAspLysGlu 96
              ::::::                          ||||||   |||      ||||||
Db       1309 AAACTTTTAGAAGAAACAAAAAATATTGACATTGAAAGTATTGAAAATTCATTAAAAGAG 1368

Qy         97 ValGluAspTyrGluLys---------AsnIleLysLysAlaArgAsnLysLeu---Arg 112
              :::|||:::   :::|||         |||::::::|||   :::   :::|||   :::
Db       1369 ATAGAGGAAAAAAAGAAAGTTCTTGAAAATCTACAAAAAGAAAAGATAGAACTAAACAAA 1428

Qy        113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
                 :::      |||      |||      ::::::      :::   :::   :::|||
Db       1429 AAACTTGGGGAAATTAACAGTGAAATTAAAAGGCTGAAAAAAATTTTAGATGAACTTAAA 1488

Qy        133 GluValGlyGlyThrLeuGlnValSerGluAsnProAspAsp 146
              ||||||   |||         :::   :::   |||   |||
```

```
Db       1489 GAAGTTGAAGGAAAATGCCCACTATGTAAAACACCAATAGAT 1530
```

RESULT 9
US-10-032-585-3062/c
; Sequence 3062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3062
; LENGTH: 55
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-3062

Alignment Scores:

Pred. No.:	0.0021	Length:	55
Score:	91.00	Matches:	18
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.79%	Indels:	0
DB:	13	Gaps:	0

US-09-980-054A-12 (1-148) x US-10-032-585-3062 (1-55)

```
Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGlu 18
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         55 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAA 2
```

RESULT 10
US-10-171-581-158
; Sequence 158, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 158
; LENGTH: 5967
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AB002371
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-158

Alignment Scores:

Pred. No.:	6.04	Length:	5967
Score:	87.00	Matches:	31
Percent Similarity:	44.97%	Conservative:	36
Best Local Similarity:	20.81%	Mismatches:	64
Query Match:	11.27%	Indels:	18
DB:	15	Gaps:	4

US-09-980-054A-12 (1-148) x US-10-171-581-158 (1-5967)

```
Qy         12 LeuGluGluGluGlnTyrGluLeuGlyPheLys------GluGlyGln------------ 25
              |||:::|||::::::      |||   :::         |||||||:::
```

```
Db       2084 CTTGATGAAAAAGAACAGGCTCTCTATTATGCTCGTTTGGAGGGAAGAAACAGAGCAAAA 2143

Qy         26 ---------------IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40

Db       2144 CATCTGCGCCAAACAATTCAGTCTCTACGACGACAGTTTAGTGGAGCTTTACCCTTGGCA 2203

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIle------------------IleGlyTyrIle 54

Db       2204 CAACAGGAAAAGTTCTCCAAAACAATGATTCAACTACAAAATGACAAACTTAAGATAATG 2263

Qy         55 GlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeu 74

Db       2264 CAAGAAATGAAAAATTCTCAACAAGAACATAGAAATATGGAGAACAAAACATTGGAGATG 2323

Qy         75 ArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAsp 94

Db       2324 GAATTAAAATTAAAGGGCCTGGAAGAGTTAATAAGCACTTTAAAGGATACCAAAGGAGCC 2383

Qy         95 LysGluValGluAspTyrGluLysAsnIleLysLysAlaArg---AsnLysLeuArgVal 113

Db       2384 CAAAAGGTAATCAACTGGCATATGAAAATAGAAGAACTTCGTCTTCAAGAACTTAAACTA 2443

Qy        114 IleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGlu 133

Db       2444 AATCGGGAATTAGTCAAGGATAAAGAAGAAATAAAATATTTGAATAACATAATTTCTGAA 2503

Qy        134 ValGlyGlyThrLeuGlnValSerGlu 142

Db       2504 TATGAACGTACAATCAGCAGTCTTGAA 2530
```

RESULT 11
US-09-932-183A-1
; Sequence 1, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7100
; TYPE: DNA
; ORGANISM: Bacillius subtilis
US-09-932-183A-1

Alignment Scores:

Pred. No.:	10.4	Length:	7100
Score:	86.00	Matches:	35
Percent Similarity:	44.23%	Conservative:	34
Best Local Similarity:	22.44%	Mismatches:	61
Query Match:	11.14%	Indels:	26
DB:	10	Gaps:	6

US-09-980-054A-12 (1-148) x US-09-932-183A-1 (1-7100)

```
Qy          5 AspIleAspAsnValLeuAsnLeu---GluGluGluGlnTyrGluLeuGlyPheLysGlu 23

Db       5677 GATGTCGATAACAAAATTTCAATGACTGAAGAAGATGAAGACAAGGTTAAATACTATAGC 5736

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43

Db       5737 AAGCAAATAAAGCTCATTCAACAACAACAAAAGGAAGCGAAGAAATACATTAAGCAGCTT 5796

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyr------------------IleGlnGluLeu 57

Db       5797 GAAGAACAAAAGAAAGCTGCGAAAGGTTTCCCTGACATCCAGGAACAGATCACTGAAGAA 5856

Qy         58 MetLysPheTrpLeuSer---------------HisIleAspGlnTyrAsnAsnSerSer 72

Db       5857 ATGCAAAACTGGAAAGATAAACAGAAAGATTTTAACCTTGAGCTTTATAACACCAAGAAG 5916

Qy         73 SerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsn 92

Db       5917 TCGATCAAGGATATCTATAAATCATTGGCTGATGAAGTTGTATCCATC------------ 5964

Qy         93 GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsn---LysLeu 111

Db       5965 --------------------TACAAAGAGATGTACGAAAAAATGCGTGATATTGAGTTA 6003

Qy        112 ArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuVal 131

Db       6004 GAAGCGCATCAGAAAGCGACTCAAGACTTG------ATCGATGAGATAGACAAGACTGAT 6057

Qy        132 LysGluValGlyGlyThrLeuGlnValSerGluAsnProAspAspMet 147

Db       6058 GACGAGGCTAAATTTCAAAAAGAATTAAAAGAAAGACAAGACAGTATT 6105
```

RESULT 12
US-09-070-927A-163/c
; Sequence 163, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 163:
US-09-070-927A-163

Alignment Scores:

```
Pred. No.:              11.1          Length:        6723
Score:                  85.50         Matches:       46
Percent Similarity:     40.00%        Conservative:  20
Best Local Similarity:  27.88%        Mismatches:    62
Query Match:            11.08%        Indels:        37
DB:                     10            Gaps:          6

US-09-980-054A-12 (1-148) x US-09-070-927A-163 (1-6723)

Qy         13 GluGluGluGln--------------TyrGluLeuGlyPheLysGluGlyGlnIleGln 27
              ||||||||||||                ||||||      ||||||   :::   |||
Db       4382 GAAGAAGAGCAAAAACGTGCAGAAGAAATTGAGTTAACAAATAAAGAAATAAAAGCCCAG 4323

Qy         28 GlyThrLysAspGlnTyrLeu---------GluGlyLysGluTyrGlyTyrGlnThrGly 44
                 |||:::  :::         |||  :::||| |||  |||   |||
Db       4322 TTTGAAGCGGATAAGCAAATCAAGCTCTCTGAAATTCGAGAAAAAGGCATTCAAAATAAC 4263

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHis 64
               |||     :::         |||  ||||||  :::         |||
Db       4262 AAAGATCGCGAAGCCCTAAAAAAAGGGATCGGCGAGTTAAAAGAACAACTATTGATTAAA 4203

Qy         65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeu-----------AsnAsn 80
              :::   ||||||  :::         :::   ||||||                |||
Db       4202 ACAGAAGCATATAACATTGCT---------AAAAAACACTTGGAAGATGTGAAGGAAAAC 4152

Qy         81 LeuGluAspIleMetAlaGlnIleSer-------------------------------- 89
              |||   |||:::     |||||||||
Db       4151 TTAGCTGATGTAGAGCAGCAGATTTCCTCATTAAAATTAGACAAAATACCATTTGAAGCA 4092

Qy         90 -----------------IleThrAsnGlyAspLysGluValGluAspTyr--------- 100
                             ||||||       |||:::::::::::|||
Db       4091 ACTGAAAAATATTCAACTATCACTAAAGAAATAAAAAAATTGCAAGAGTACATTACTCAA 4032

Qy        101 ------GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThr 118
                    |||     :::::::::         |||     |||     |||
Db       4031 AGTAACGAAGCTATTCTAGAACAAACTTCCGCTAAAACAAGTGAGATTACAGAAATAGAT 3972

Qy        119 LysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeu 138
              ||||||      ||||||           |||:::||||||                |||
Db       3971 AAAGAAATCGCAATGATTGATGAAAAATTAGCCTTGCTAAAAGAGTATGAAAGACAGCTA 3912

Qy        139 GlnValSerGluAsn 143
                 :::   |||:::
Db       3911 TCGATTATTGAAGAT 3897

RESULT 13
US-10-369-493-44827
; Sequence 44827, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44827
;   LENGTH: 255
;   TYPE: DNA
;   ORGANISM: Xenorhabdus nematophilus
US-10-369-493-44827

Alignment Scores:
Pred. No.:              0.111         Length:        255
Score:                  85.00         Matches:       28
Percent Similarity:     48.24%        Conservative:  13
Best Local Similarity:  32.94%        Mismatches:    34
Query Match:            11.01%        Indels:        10
DB:                     12            Gaps:          4

US-09-980-054A-12 (1-148) x US-10-369-493-44827 (1-255)

Qy         65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
              :::   |||       ||||||       ||||||       |||   :::|||   |||
Db          1 ATGTCGCAGAAGAAAAACTCTGAAGATTTACGTACTGAATTGCAATCCCTTGCGGATACG 60

Qy         85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
              :::   ::::::      ::::::||||||||          :::   ||||||
Db         61 CTGGAAGAAGTTCTCAATAGTTCCGGTGACAAATCCAAGGCTGAATTGGAAAAACTACGT 120

Qy        105 LysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIle 124
               ||||||       |||:::       |||   :::|||
Db         121 AGTAAAGCAGAAAAAACGTTGAAA---GATGCGCGAGTCACGCTCAGT------------ 165

Qy        125 AspSerLeuAspAsnLeuVal---------LysGluValGlyGlyThrLeuGln------ 139
              ||:::   |||   ||||||          ||||||:::   |||
Db         166 GATGCCAGCGATAAGTTGGTCGATCAAACAAAGGAAATGGCAGGTCGTGCGGATAATTAT 225

Qy        140 ValSerGluAsnPro 144
              |||   :::|||||
Db         226 GTTCGTGACAATCCG 240

RESULT 14
US-10-369-493-44024
; Sequence 44024, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44024
;   LENGTH: 2637
;   TYPE: DNA
;   ORGANISM: Pyrococcus horikoshii
US-10-369-493-44024

Alignment Scores:
Pred. No.:              5.09          Length:        2637
Score:                  83.50         Matches:       37
Percent Similarity:     41.18%        Conservative:  40
Best Local Similarity:  19.79%        Mismatches:    61
Query Match:            10.82%        Indels:        49
DB:                     12            Gaps:          6

US-09-980-054A-12 (1-148) x US-10-369-493-44024 (1-2637)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              :::|||   |||:::::::::::::   |||       ::::::|||:::|||||||||
Db       1534 TTAAGCAAGATAAATCTGGAAGAT---TTGAAGAGGGATAAGGAAGAGTACGAGTTACTA 1590

Qy         21 PheLysGlu--------------GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGlu 35
                 |||                |||::::::::::     |||   :::       |||
Db       1591 AAAAGTGAGAGCAATAAATTAAAGGGTGAAGTGGAGAGCCTAAAAAAGGAAGTAAATGAG 1650

Qy         36 GlyLysGluTyrGlyTyrGlnThr---------------------------------- 43
                 :::|||      ::::::
Db       1651 CTTAATGATTATAAAAATGAATCTACCAAGCTTGAAATCGAAATAGATAAGGCCAAAAAG 1710

Qy         44 ---------------------------------GlyPheGlnArgPhe---LeuIle 50
```

```
                                                 ||||||:::             :::
Db       1711 GAATTATCTGAGATCGAAGATAGGTTACTAAGGTTGGGATTTAAAACAATAGATGAGCTT 1770

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
              |||   |||:::||||||  ||||||         :::|||:::      |||
Db       1771 TCTGGGAGGATAAGGGAACTTGAAAAGTTTCACAACAAATACATAGAAGCAAAAAACGCT 1830

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle------ 88
                 ||||||:::   |||  :::|||:::|||         ::::::
Db       1831 GAGAAGGAGTTAAGGGACATCCTTGAAAGTCTTAAAGACGAGAGGGAGGAACTTGATAAG 1890

Qy         89 -------------------------------------SerIleThrAsnGlyAspLys 95
                                                   :::|||:::
Db       1891 GCCTTCGAAGAACTAGCAAAAATTGAAACGGATATAGAAAAAGTCACTTCACAGCTTAAC 1950

Qy         96 GluValGlu---------AspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArg 112
              |||::::::         |||  :::|||     :::|||            ::::::
Db       1951 GAACTTCAAAGGAAATTTGACCAAAAGAAGTACGAAGAAAAAAGGGAGAAGATGATGAAG 2010

Qy        113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
              :::      |||                |||::::::   |||:::
Db       2011 CTTAGTATGGAGATAAAGGGATTGGAAACTAAGTTAGAGGAACTTGAGAGAAGAAGGGAT 2070

Qy        133 GluValGlyGlyThrLeuGln 139
              |||:::     |||::::::
Db       2071 GAAATTAAATCTACAATAGAA 2091

RESULT 15
US-10-369-493-24958
; Sequence 24958, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24958
;   LENGTH: 2640
;   TYPE: DNA
;   ORGANISM: Pyrococcus horikoshii
US-10-369-493-24958

Alignment Scores:
Pred. No.:              5.1          Length:          2640
Score:                  83.50        Matches:         37
Percent Similarity:     41.18%       Conservative:    40
Best Local Similarity:  19.79%       Mismatches:      61
Query Match:            10.82%       Indels:          49
DB:                     12           Gaps:            6

US-09-980-054A-12 (1-148) x US-10-369-493-24958 (1-2640)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              :::|||   |||:::::::::::   |||      :::::::|||:::|||||||||
Db       1534 TTAAGCAAGATAAATCTGGAAGAT---TTGAAGAGGGATAAGGAAGAGTACGAGTTACTA 1590

Qy         21 PheLysGlu---------------GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGlu 35
                 |||               |||:::::::::      |||   :::     |||
Db       1591 AAAAGTGAGAGCAATAAATTAAAGGGTGAAGTGGAGAGCCTAAAAAAGGAAGTAAATGAG 1650

Qy         36 GlyLysGluTyrGlyTyrGlnThr---------------------------------- 43
                :::|||      ::::::
Db       1651 CTTAATGATTATAAAAATGAATCTACCAAGCTTGAAATCGAAATAGATAAGGCCAAAAAG 1710

Qy         44 ----------------------------------GlyPheGlnArgPhe---LeuIle 50
                                                ||||||:::           :::
Db       1711 GAATTATCTGAGATCGAAGATAGGTTACTAAGGTTGGGATTTAAAACAATAGATGAGCTT 1770

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
              |||   |||:::||||||  ||||||         :::|||:::      |||
Db       1771 TCTGGGAGGATAAGGGAACTTGAAAAGTTTCACAACAAATACATAGAAGCAAAAAACGCT 1830

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle------ 88
                 ||||||:::   |||  :::|||:::|||         ::::::
Db       1831 GAGAAGGAGTTAAGGGACATCCTTGAAAGTCTTAAAGACGAGAGGGAGGAACTTGATAAG 1890

Qy         89 -------------------------------------SerIleThrAsnGlyAspLys 95
                                                   :::|||:::
Db       1891 GCCTTCGAAGAACTAGCAAAAATTGAAACGGATATAGAAAAAGTCACTTCACAGCTTAAC 1950

Qy         96 GluValGlu---------AspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArg 112
              |||::::::         |||  :::|||     :::|||            ::::::
Db       1951 GAACTTCAAAGGAAATTTGACCAAAAGAAGTACGAAGAAAAAAGGGAGAAGATGATGAAG 2010

Qy        113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
              :::      |||                |||::::::   |||:::
Db       2011 CTTAGTATGGAGATAAAGGGATTGGAAACTAAGTTAGAGGAACTTGAGAGAAGAAGGGAT 2070

Qy        133 GluValGlyGlyThrLeuGln 139
              |||:::     |||::::::
Db       2071 GAAATTAAATCTACAATAGAA 2091

Search completed: February 16, 2004, 17:23:44
Job time : 352 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:25:48 ; Search time 278 Seconds
(without alignments)
1437.111 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980054/runat_13022004_163838_26290/app_query.fasta_1.327
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054_@CGN_1_1_312_@runat_13022004_163838_26290 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	772	100.0	447	22	AAC86750	DNA encoding a Can
	2	764	99.0	447	22	AAS23431	Candida albicans e
	3	764	99.0	447	24	ABZ31775	Candida albicans e
	4	562	72.8	326	22	AAC86756	Probe for DNA enco
c	5	119.5	15.5	2877	25	ABT20223	Aspergillus fumiga
c	6	119.5	15.5	3863	25	ABT17813	Aspergillus fumiga
c	7	119.5	15.5	4877	25	ABT19627	Aspergillus fumiga
	8	116	15.0	1372	21	AAC42626	Arabidopsis thalia
	9	107.5	13.9	2462	20	AAX20299	Borrelia burgdorfe
c	10	107	13.9	65	24	ABZ28869	Candida gene relat
c	11	106.5	13.8	2601	25	ABT20821	Aspergillus fumiga
	12	103.5	13.4	1131	22	ABA09167	Human GK003 homolo
	13	103.5	13.4	1131	22	AAK52584	Human polynucleoti
	14	99.5	12.9	445	21	AAC01972	Human secreted pro
	15	99.5	12.9	677	24	ABQ56851	Human colon cancer
	16	99.5	12.9	1118	22	AAK51600	Human polynucleoti
	17	99.5	12.9	1164	23	AAS92292	DNA encoding novel
c	18	97.5	12.6	111309	20	AAX20250	Borrelia burgdorfe
c	19	97.5	12.6	910715	20	AAX20248	Borrelia burgdorfe
	20	95.5	12.4	426	23	ABL22081	Drosophila melanog
	21	95.5	12.4	2306	23	ABL22078	Drosophila melanog
	22	95.5	12.4	2426	23	ABL22080	Drosophila melanog
	23	95.5	12.4	3534	23	ABL22082	Drosophila melanog
	24	95	12.3	7057	24	ABZ11230	Human polynucleoti
	25	94.5	12.2	2991	20	AAX20288	Borrelia burgdorfe
	26	91.5	11.9	1664976	19	AAV21209	Methanococcus jann
c	27	91	11.8	55	24	ABZ28979	Candida gene relat
c	28	91	11.8	80	22	AAS23676	Tetracycline promo
	29	90.5	11.7	48551	24	AAS20800	Clostridium diffic
	30	90	11.7	1428	23	ABL23453	Drosophila melanog
c	31	90	11.7	35515	20	AAX20252	Borrelia burgdorfe
	32	89	11.5	4228	20	AAX20277	Borrelia burgdorfe
	33	88.5	11.5	2422	23	ABL14444	Drosophila melanog
	34	87	11.3	717	20	AAX61834	B. burgdorferi ant
	35	87	11.3	828	20	AAX61833	B. burgdorferi ant
	36	87	11.3	53585	20	AAX20251	Borrelia burgdorfe
	37	87	11.3	487980	25	AAD53223	Human chromosome 3
c	38	86.5	11.2	1902	20	AAX20309	Borrelia burgdorfe
	39	86	11.1	4731	23	ABL23452	Drosophila melanog
	40	86	11.1	7100	20	AAX24980	Bacillus subtilis
	41	85.5	11.1	657	24	ABQ58583	Human colon cancer
c	42	85.5	11.1	6723	20	AAX13100	Enterococcus faeca
c	43	85.5	11.1	6723	24	ABS98895	Enterococcus faeca
	44	85	11.0	840	20	AAX20354	Borrelia burgdorfe
	45	85	11.0	1524	21	AAA70107	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AAC86750
ID AAC86750 standard; DNA; 447 BP.
XX
AC AAC86750;
XX
DT 02-APR-2001 (first entry)
XX
DE DNA encoding a Candida albicans protein CaNL260.
XX
KW CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CaNL260; CaDR361;
KW antifungal; fungal infection; pathogenic fungi; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 1..447

```
FT                   /*tag=  a
XX
PN   WO200075305-A2.
XX
PD   14-DEC-2000.
XX
PF   08-JUN-2000; 2000WO-FR01567.
XX
PR   09-JUN-1999;   99FR-0007250.
XX
PA   (HMRI ) HOECHST MARION ROUSSEL.
XX
PI   Lalanne J,  Rocher C;
XX
DR   WPI; 2001-050024/06.
DR   P-PSDB; AAB30859.
XX
PT   New polynucleotides from Candida albicans and their derived proteins,
PT   useful for diagnosis and treatment of fungal infections and for drug
PT   screening  -
XX
PS   Claim 4; Page 77-78; 89pp; French.
XX
CC   The present sequence encodes a Candida albicans protein. The
CC   specification describes genes CaDRA472, CaDR489, 1CaDR527, 2CaDR527,
CC   CaFLO24, CaNL260, and CaDR361. These genes are essential for
CC   survival, and so are good targets for antifungal agents. The Candida
CC   albicans genes and their derived proteins are used to screen compounds
CC   for the ability to inhibit the activity of the protein, i.e. for
CC   antifungal activity. The proteins are also used to generate a protective
CC   antibody response against fungal infections in mammals. The Candida
CC   albicans proteins and genes, and their antibodies, are used for
CC   diagnosing fungal infections, specifically C. albicans (in standard
CC   amplification, hybridisation or immunological assays, and for studying
CC   pathogenic fungi.
XX
SQ   Sequence 447 BP; 192 A; 38 C; 82 G; 135 T; 0 other;

Alignment Scores:
Pred. No.:              1.79e-80        Length:         447
Score:                  772.00          Matches:        148
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     22              Gaps:           0

US-09-980-054A-12 (1-148) x AAC86750 (1-447)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        141 SerGluAsnProAspAspMetTrp 148
              |||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGG 444

RESULT 2
AAS23431
ID   AAS23431 standard; DNA; 447 BP.
XX
AC   AAS23431;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Candida albicans essential gene CaYNL260C.
XX
KW   Gene identification; essential gene; GRACE; pathogenic fungus;
KW   gene replacement and conditional expression; fungal infection; ds.
XX
OS   Candida albicans.
XX
PN   WO200160975-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05551.
XX
PR   18-FEB-2000; 2000US-0183534.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H;
XX
DR   WPI; 2001-489080/53.
DR   P-PSDB; AAU15103.
XX
PT   Identifying genes essential to fungal metabolisms and identifying
PT   potential therapeutic agents that target these genes -
XX
PS   Claim 22; Page 167; 324pp; English.
XX
CC   The present invention relates to novel methods for constructing fungal
CC   strains useful for identification and validation of gene products as
CC   targets for therapeutic agents, for creating a collection of identified
CC   essential genes, and screening assays for the discovery of new drugs.
CC   The invention provides the GRACE (gene replacement and conditional
CC   expression) method for the construction of mutant organisms referred to
CC   as GRACE strains of the organism. The invention can be applied to any
CC   organism, particularly a pathogenic fungus e.g. Candida albicans,
CC   Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC   useful to identify agents that may be used in the treatment of fungal
CC   infections. AAS23381-AAS23442 represent C. albicans essential genes.
XX
SQ   Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Alignment Scores:
Pred. No.:              1.51e-79        Length:         447
Score:                  764.00          Matches:        146
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  98.65%          Mismatches:     0
Query Match:            98.96%          Indels:         0
DB:                     22              Gaps:           0

US-09-980-054A-12 (1-148) x AAS23431 (1-447)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
```

```
Qy          41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy          61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy          81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
               ||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy         101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy         121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy         141 SerGluAsnProAspAspMetTrp 148
               |||||||||||||||||||||||
Db         421 AGTGAAAACCCCGATGATATGTGG 444
```

RESULT 3
ABZ31775
ID ABZ31775 standard; DNA; 447 BP.
XX
AC ABZ31775;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6062.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49486.
XX
PR 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
DR P-PSDB; ABP73225.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression -
XX
PS Claim 37; SEQ ID NO 6062; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Alignment Scores:

Pred. No.:	1.51e-79	Length:	447
Score:	764.00	Matches:	146
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	98.65%	Mismatches:	0
Query Match:	98.96%	Indels:	0
DB:	24	Gaps:	0

US-09-980-054A-12 (1-148) x ABZ31775 (1-447)

```
Qy           1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy          21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy          41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy          61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy          81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
               ||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy         101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy         121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy         141 SerGluAsnProAspAspMetTrp 148
               |||||||||||||||||||||||
Db         421 AGTGAAAACCCCGATGATATGTGG 444
```

RESULT 4
AAC86756
ID AAC86756 standard; DNA; 326 BP.
XX
AC AAC86756;
XX
DT 02-APR-2001 (first entry)
XX
DE Probe for DNA encoding a Candida albicans protein CaNL260.
XX
KW CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CaNL260; CaDR361;

KW antifungal; fungal infection; pathogenic fungi; probe; ss.
XX
OS Candida albicans.
XX
PN WO200075305-A2.
XX
PD 14-DEC-2000.
XX
PF 08-JUN-2000; 2000WO-FR01567.
XX
PR 09-JUN-1999; 99FR-0007250.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Lalanne J, Rocher C;
XX
DR WPI; 2001-050024/06.
XX
PT New polynucleotides from Candida albicans and their derived proteins,
PT useful for diagnosis and treatment of fungal infections and for drug
PT screening -
XX
PS Example 5; Page 85-86; 89pp; French.
XX
CC The present sequence represents a probe for DNA encoding a Candida
CC albicans protein. The specification describes genes CaDRA472, CaDR489,
CC 1CaDR527, 2CaDR527, CaFLO24, CaNL260, and CaDR361. These genes are
CC essential for survival, and so are good targets for antifungal agents.
CC The Candida albicans genes and their derived proteins are used to screen
CC compounds for the ability to inhibit the activity of the protein,
CC i.e. for antifungal activity. The proteins are also used to generate a
CC protective antibody response against fungal infections in mammals. The
CC Candida albicans proteins and genes, and their antibodies, are used for
CC diagnosing fungal infections, specifically C. albicans (in standard
CC amplification, hybridisation or immunological assays, and for studying
CC pathogenic fungi.
XX
SQ Sequence 326 BP; 146 A; 27 C; 53 G; 100 T; 0 other;

Alignment Scores:

Pred. No.:	2.68e-56	Length:	326
Score:	562.00	Matches:	107
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.07%	Mismatches:	0
Query Match:	72.80%	Indels:	0
DB:	22	Gaps:	0

US-09-980-054A-12 (1-148) x AAC86756 (1-326)

```
Qy          5 AspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGly 24
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db          2 GATATAGATAATGTATTAAATTTAGAAGAAGATCAATATGAATTAGGATTTAAAGAAGGT 61

Qy         25 GlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGA 121

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHis 64
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCAT 181

Qy         65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATT 241

Qy         85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATT 301

Qy        105 LysLysAlaArgAsnLysLeuArg 112
              |||||||||||||||||||||||
Db        302 AAAAAGGCAAGAAATAAATTAAGA 325
```

RESULT 5
ABT20223/c
ID ABT20223 standard; DNA; 2877 BP.
XX
AC ABT20223;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene #2581.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention.
XX
SQ Sequence 2877 BP; 732 A; 751 C; 744 G; 650 T; 0 other;

Alignment Scores:

Pred. No.:	0.000849	Length:	2877
Score:	119.50	Matches:	34
Percent Similarity:	40.67%	Conservative:	27

```
Best Local Similarity:  22.67%        Mismatches:    48
Query Match:            15.48%        Indels:        41
DB:                     25            Gaps:          3

US-09-980-054A-12 (1-148) x ABT20223 (1-2877)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::||||||||||||   |||:::   ||||:::     |||
Db        539 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 480

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::     |||   ||||||      :::     :::    ||||||
Db        479 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 420

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||   :::|||   :::                  |||         :::
Db        419 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 360

Qy         65 ------------------------------------------------------------ 65

Db        359 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 300

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                  ||||||
Db        299 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 240

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                    :::||||||   |||      :::|||       |||   :::::::::
Db        239 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 180

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                |||   :::   :::   |||   ::::::   |||   |||   ::::::   :::
Db        179 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 120

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
              |||   |||       |||   ::::::|||
Db        119 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 90

RESULT 6
ABT17813/c
ID   ABT17813 standard; DNA; 3863 BP.
XX
AC   ABT17813;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #171.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer  -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention.
XX
SQ   Sequence 3863 BP; 983 A; 980 C; 970 G; 930 T; 0 other;

Alignment Scores:
Pred. No.:              0.00123       Length:        3863
Score:                  119.50        Matches:       34
Percent Similarity:     40.67%        Conservative:  27
Best Local Similarity:  22.67%        Mismatches:    48
Query Match:            15.48%        Indels:        41
DB:                     25            Gaps:          3

US-09-980-054A-12 (1-148) x ABT17813 (1-3863)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::||||||||||||   |||:::   |||:::     |||
Db        677 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 618

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||   :::     |||   ||||||      :::     :::    ||||||
Db        617 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 558

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||   :::|||   :::                  |||         :::
Db        557 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 498

Qy         65 ------------------------------------------------------------ 65

Db        497 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 438

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                  ||||||
Db        437 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 378

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                    :::||||||   |||      :::|||       |||   :::::::::
Db        377 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 318

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                |||   :::   :::   |||   ::::::   |||   |||   ::::::   :::
Db        317 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 258

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
```

```
                ||| |||     |||  ::::::|||
Db        257 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 228

RESULT 7
ABT19627/c
ID   ABT19627 standard; DNA; 4877 BP.
XX
AC   ABT19627;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene #1985.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US13142.
XX
PR   23-APR-2001; 2001US-285697P.
PR   27-APR-2001; 2001US-287066P.
PR   05-JUN-2001; 2001US-295890P.
PR   09-JUL-2001; 2001US-303899P.
PR   31-AUG-2001; 2001US-316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of
PT   Aspergillus fumigatus, useful for treating or preventing infections by
PT   A. fumigatus, or for treating a non-infectious disease in a subject
PT   e.g. cancer -
XX
PS   Disclosure; Page -; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention.
XX
SQ   Sequence 4877 BP; 1260 A; 1211 C; 1252 G; 1154 T; 0 other;

Alignment Scores:

Pred. No.:                  0.00166     Length:          4877
Score:                      119.50      Matches:         34
Percent Similarity:         40.67%      Conservative:    27
Best Local Similarity:      22.67%      Mismatches:      48
Query Match:                15.48%      Indels:          41
DB:                         25          Gaps:            3

US-09-980-054A-12 (1-148) x ABT19627 (1-4877)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::     |||
Db       1539 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 1480

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                |||  :::     |||  ||||||     :::     :::  ||||||
Db       1479 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 1420

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              ::::::||||||  :::|||  :::             |||         :::
Db       1419 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 1360

Qy         65 ------------------------------------------------------------ 65

Db       1359 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGACGGAACAGATACCCGCAAA 1300

Qy         66 ---AspGlnTyrAsn---------------------------------------------- 69
                  ||||||
Db       1299 TCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCTG 1240

Qy         70 ---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
                     :::||||||  |||    :::|||     |||  :::::::::
Db       1239 CCTCTCCCCTCCAGTTCACGGCTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGAT 1180

Qy         86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
                |||  :::  :::  |||  ::::::  |||  |||  ::::::  :::
Db       1179 CCAGCTTCGTTGCCTATGGAGAACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTT 1120

Qy        105 LysLysAlaArgAsnLysLeuArgValIle 114
              |||  |||     |||  ::::::|||
Db       1119 AAGGGCGCTGCAATCAAGGCCAAGCTTATC 1090

RESULT 8
AAC42626
ID   AAC42626 standard; DNA; 1372 BP.
XX
AC   AAC42626;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 36260.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
```

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Alignment Scores:
Pred. No.:              0.000848        Length:         1372
Score:                  116.00          Matches:        32
Percent Similarity:     43.90%          Conservative:   22
Best Local Similarity:  26.02%          Mismatches:     51
Query Match:            15.03%          Indels:         18
DB:                     21              Gaps:           2

US-09-980-054A-12 (1-148) x AAC42626 (1-1372)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   ::::::   |||||||||         :::   ||||||   ||||||
Db        185 CTCGATTGTATCGTGCGTTTAGAGGAGACACATGTTCAACAAGGCTTCGATGAGGGTTAC 244

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                ::|||                 :::  :::      |||  :::   ||||||
Db        245 GAAGAAGGTCTTGTGTCCGGTCGTGAAGACGCTCGTCATCTGGGTTTGAAACTCGGGTTC 304

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------ 63
              :::         :::||||||:::   :::              |||   |||
Db        305 GAGACAGGCGAGCTGATTGGATTCTACAGAGGTTGCTCTGCTCTTTGGAATTCAGCTCTC 364

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                ||||||          |||     |||      |||||||||:::
Db        365 CGTATTGATCCTACACGCTTCTCTCCTCAGCTCCATAAGCATCTCAATGATTTCCATGTC 424

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              ::::::   :::|||   :::   :::   :::   |||   :::|||       |||:::
Db        425 TTGCTCGATAAAATCCCGCTTTTGGATCCCGAGGACGAAGCTAAAGACGGGATCAAGGAT 484

Qy        104 ------------------------------------------------IleLysLysAla 107
                                                              |||   :::|||
Db        485 GATCTCAGAGAGATGTTGAGAGAAATGCTTGGTTGTTGCAAGGTTTACATATCTGAAGCG 544

Qy        108 ArgAsnLys 110
                ||||||||
Db        545 CGGAATAAG 553

RESULT 9
AAX20299
ID   AAX20299 standard; DNA; 2462 BP.
XX
AC   AAX20299;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #52.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 1030-1031; 1128pp; English.
XX
CC   AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 2462 BP; 1080 A; 229 C; 357 G; 792 T; 4 other;

Alignment Scores:
Pred. No.:              0.0172          Length:         2462
Score:                  107.50          Matches:        46
Percent Similarity:     43.21%          Conservative:   24
Best Local Similarity:  28.40%          Mismatches:     55
Query Match:            13.92%          Indels:         37
DB:                     20              Gaps:           9

US-09-980-054A-12 (1-148) x AAX20299 (1-2462)

Qy         13 GluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGly---ThrLysAsp 31
              ::::::|||   |||:::         |||   |||:::      |||         :::
Db       1521 AAGCAAGAAAGTTATGATTTTTATAAAAAAAATGGGAAGTTTACTGGGTTTTTATTGGAA 1580

Qy         32 GlnTyrLeuGluGly----------------------------LysGluTyr--------- 39
              :::   ||||||                             |||:::|||
Db       1581 GAACTTCTTGAAAGTCAAACAGATTTGATTAATAAGCTTTTAAAAAAATATAAACAATTA 1640

Qy         40 ---GlyTyrGlnThrGlyPheGlnArg-------------------------------- 47
                  ||||||         |||  |||
Db       1641 AAAGGATATTAAGAGGATTTTATGAGAAATTTGGTGCACAGAACATATGATATAGAAAGC 1700

Qy         48 ------------PheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
                         ||||||   ||||||:::   :::|||   :::   |||
Db       1701 ATAAAAAATGAATTTTTAAACATAGGATTTAGTGAAGAGGCAATAGATTTTGTTTTTCTT 1760

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
              |||   |||   |||   |||          |||:::      :::   ::::::|||
Db       1761 CATAATGATAATTAC---AACTATGAGGTTTTAAAAGAGAAAATAATTGATGTAGAAAAG 1817
```

```
Qy          84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
               :::   |||      |||::::::   |||   ::::::::::::   ||||||||||
Db        1818 AATTTG---CAAAAAGACATATCTAGTTTAGATACTAAGATAGATAATGTAGAAAAGAAT 1874

Qy         104 ------------IleLysLysAlaArgAsnLysLeuArgVal---IleAlaSerIleThr 118
                        |||                        |||   :::   |||   |||:::
Db        1875 TTAAATGTTAAGATAGATAATGTAGAGAAGAATTTAAATATTAAAATAGATAGTGTTAAA 1934

Qy         119 LysGlu---ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThr 137
                |||          ||||||||||||||||||      :::      |||      |||
Db        1935 AATGAACTTAATTCTAAAATAGATAGTTTAGATACTAAGATAGATAATGTAGAAAAAACT 1994

Qy         138 LeuGln 139
               ||||||
Db        1995 TTGCAA 2000
```

RESULT 10
ABZ28869/c
ID ABZ28869 standard; DNA; 65 BP.
XX
AC ABZ28869;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida gene related tetracyclin promoter PCR primer SEQ ID NO 2884.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49486.
XX
PR 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression -
XX
PS Claim 76; SEQ ID NO 2884; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of a PCR
CC primer used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 65 BP; 20 A; 11 C; 2 G; 32 T; 0 other;

Alignment Scores:

Pred. No.:	0.000198	Length:	65
Score:	107.00	Matches:	21
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	13.86%	Indels:	0
DB:	24	Gaps:	0

US-09-980-054A-12 (1-148) x ABZ28869 (1-65)

```
Qy           2 SerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPhe 21
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          65 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 6

Qy          22 Lys 22
               |||
Db           5 AAA 3
```

RESULT 11
ABT20821/c
ID ABT20821 standard; DNA; 2601 BP.
XX
AC ABT20821;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene #3179.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object

```
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This polynucleotide sequence represents one of the essential
CC   genes of Aspergillus fumigatus of the invention.
XX
SQ   Sequence 2601 BP; 671 A; 675 C; 677 G; 578 T; 0 other;

Alignment Scores:
Pred. No.:                0.024          Length:          2601
Score:                    106.50         Matches:         35
Percent Similarity:       42.66%         Conservative:    26
Best Local Similarity:    24.48%         Mismatches:      48
Query Match:              13.80%         Indels:          35
DB:                       25             Gaps:            3

US-09-980-054A-12 (1-148) x ABT20821 (1-2601)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   :::|||:::|||||||||||||   |||:::   |||:::      |||
Db        517 CTTGACGGTCTTCTCGACCTGGAAGAGGAATTCTACCAAGAGGGCTATAACTTGGGTGCA 458

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                |||   :::      |||   ||||||      :::      :::   ||||||
Db        457 ACTGACGGTGCCCAGGCTGGTTACACTGAAGGAAGCGTGTTTGCCGTTGAGAAAGGCTTC 398

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeu--------- 62
              ::::::||||||   :::|||   :::             |||
Db        397 GAAAAGTTTCTTGAATTGGGAAGACTATATGGCAAAGCACTCGTCTGGGCCCAGAGGCTC 338

Qy         63 ------SerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHis--------- 77
                    |||   |||   :::   ::::::      |||   ||||||
Db        337 GCAGACTCTGAAATCTCTCGGAATTCGTCACAGGAGAGTGA-CGGAACAGATACCCGCAA 279

Qy         77 ------------------------------------------------------------ 77

Db        278 ATCTCAGCCATATAATGAGGGTGATCTTCTCTTGGAACCATTCGTATGCAAGGAGATGCG 219

Qy         78 ---LeuAsnAsnLeuGluAspIleMet--------------AlaGlnIleSerIleThr 91
                  |||||||||   ::::::              |||   :::   :::
Db        218 CTGGCCAAAAACCTCGAGATTCTACTCGAATTAGTTGATCCAGCTTCGTTGCCTATGGAG 159

Qy         92 AsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeu 111
              |||   ::::::   |||   |||   ::::::   :::|||   |||      |||
Db        158 AACACGGAAGAAGCTGTCACAGATGTTGATGAGCGTCTTAAGGGCGCTGCAATCAAGGCC 99

Qy        112 ArgValIle 114
              ::::::|||
Db         98 AAGCTTATC 90

RESULT 12
ABA09167
ID   ABA09167 standard; cDNA; 1131 BP.
XX
AC   ABA09167;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human GK003 homologue-encoding cDNA, SEQ ID NO:943.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   P-PSDB; ABB11923.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 1; Page 809; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
```

```
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a cDNA encoding a
CC   novel human polypeptide of the invention.
XX
SQ   Sequence 1131 BP; 340 A; 266 C; 253 G; 270 T; 2 other;

Alignment Scores:
Pred. No.:              0.0187          Length:         1131
Score:                  103.50          Matches:        43
Percent Similarity:     44.38%          Conservative:   32
Best Local Similarity:  25.44%          Mismatches:     60
Query Match:            13.41%          Indels:         34
DB:                     22              Gaps:           7

US-09-980-054A-12 (1-148) x ABA09167 (1-1131)

Qy         11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
              |||::::::          |||   |||::::::::::|||       |||
Db        361 AACATGCAAAGACGAGTCAAAGAA---GGTTATAGAGATGGAATAGATGCTGGCAAAGCA 417

Qy         31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 50
                    :::|||          ||||||:::   |||   :::       |||
Db        418 GTTACTCTTCAACAGGGCTTCAATCAAGGTTATAAGAAAGGTGCAGAAGTCATTTTAAAC 477

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
                |||   ::::::      :::          ||||||              :::||||||
Db        478 TATGGACGACTCCGAGGAACATTGAGTGCTTTGCTCTCCTGGTGTCACCTTCATAATAAT 537

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle------ 88
              :::|||:::|||   |||   :::|||||||||   |||   :::   |||
Db        538 AATTCAACTTTGATCAATAAAATAAACAATCTTCTGGATGCAGTTGGCCAGTGTGAAGAG 597

Qy         89 ----------------------SerIleThrAsnGlyAsp-----------------Lys 95
                                    |||||||||
Db        598 TATGTGCTCAAACATCTGAAATCAATCACTCCACCGTCCCATGTTGTAGATTTATTGGAC 657

Qy         96 GluValGluAspTyr-----------------------GluLysAsnIleLysLysAla 107
                :::||||||                         ||||||   |||   :::|||
Db        658 TCCATTGAGGATATGGACCTTTGTCATGTAGTTCCAGCTGAGAAAAAGATTGATGAAGCT 717

Qy        108 ArgAsn-Lys------LeuArgValIleAlaSerIleThrLysGluThrTrpLys----- 123
              :::::: |||      ::::::::::::   |||:::||||||         :::
Db        718 AAAGATGAAAGACTCTGTGAAAATAATGCTGAGTTTAACAAAAACTGTAGCAAGAGCCAT 777

Qy        124 ----------IleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVa 140
                       |||      :::   |||:::|||:::          ::::::   ::
Db        778 AGTGGGATAGATTGTTCATATGTAGAATGTTGTAGAACAC------AGGAGCATGCACAT 831

Qy        140 lSerGluAsnProAspAspMetTrp 148
              :  |||||||||        |||
Db        832 TCAGGAAAACCCAAGCCCCACATGG 856

RESULT 13
AAK52584
ID   AAK52584 standard; cDNA; 1131 BP.
XX
AC   AAK52584;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2113.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79451.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 4491; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence 1131 BP; 340 A; 266 C; 253 G; 270 T; 2 other;

Alignment Scores:
Pred. No.:              0.0187          Length:         1131
Score:                  103.50          Matches:        43
Percent Similarity:     44.38%          Conservative:   32
Best Local Similarity:  25.44%          Mismatches:     60
Query Match:            13.41%          Indels:         34
DB:                     22              Gaps:           7

US-09-980-054A-12 (1-148) x AAK52584 (1-1131)

Qy         11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
              |||::::::          |||   |||::::::::::|||       |||
Db        361 AACATGCAAAGACGAGTCAAAGAA---GGTTATAGAGATGGAATAGATGCTGGCAAAGCA 417

Qy         31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 50
                    :::|||          ||||||:::   |||   :::       |||
Db        418 GTTACTCTTCAACAGGGCTTCAATCAAGGTTATAAGAAAGGTGCAGAAGTCATTTTAAAC 477

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
                |||   ::::::      :::          ||||||              :::||||||
Db        478 TATGGACGACTCCGAGGAACATTGAGTGCTTTGCTCTCCTGGTGTCACCTTCATAATAAT 537

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle------ 88
              :::|||:::|||   |||   :::|||||||||   |||   :::   |||
Db        538 AATTCAACTTTGATCAATAAAATAAACAATCTTCTGGATGCAGTTGGCCAGTGTGAAGAG 597

Qy         89 ----------------------SerIleThrAsnGlyAsp-----------------Lys 95
                                    |||||||||
Db        598 TATGTGCTCAAACATCTGAAATCAATCACTCCACCGTCCCATGTTGTAGATTTATTGGAC 657
```

```
Qy         96 GluValGluAspTyr----------------------GluLysAsnIleLysLysAla 107
              :::||||||                          ||||||   |||   :::|||
Db        658 TCCATTGAGGATATGGACCTTTGTCATGTAGTTCCAGCTGAGAAAAAGATTGATGAAGCT 717

Qy        108 ArgAsn-Lys------LeuArgValIleAlaSerIleThrLysGluThrTrpLys----- 123
              :::::: |||      ::::::::::::   |||:::||||||         :::
Db        718 AAAGATGAAAGACTCTGTGAAAATAATGCTGAGTTTAACAAAAACTGTAGCAAGAGCCAT 777

Qy        124 ----------IleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVa 140
                     |||      :::   |||:::|||:::           ::::::   ::
Db        778 AGTGGGATAGATTGTTCATATGTAGAATGTTGTAGAACAC------AGGAGCATGCACAT 831

Qy        140 lSerGluAsnProAspAspMetTrp 148
              :  |||||||||        |||
Db        832 TCAGGAAAACCCAAGCCCCACATGG 856
```

RESULT 14
AAC01972
ID AAC01972 standard; cDNA; 445 BP.
XX
AC AAC01972;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1970.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG01966.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1970; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 445 BP; 132 A; 86 C; 111 G; 114 T; 2 other;

Alignment Scores:
Pred. No.: 0.0167 Length: 445
Score: 99.50 Matches: 32
Percent Similarity: 45.83% Conservative: 23
Best Local Similarity: 26.67% Mismatches: 46
Query Match: 12.89% Indels: 19
DB: 21 Gaps: 3

US-09-980-054A-12 (1-148) x AAC01972 (1-445)

```
Qy         11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
              |||::::::             |||   |||:::::::::|||          |||
Db        121 AACATGCAAAGACGAGTCAAAGAA---GGTTATAGAGATGGAATAGATGCTGGCAAAGCA 177

Qy         31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 50
                     :::|||          ||||||:::   |||   :::      |||
Db        178 GTTACTCTTCAACAGGGCTTCAATCAAGGTTATAAGAAAGGTGCAGAAGTCATTTTAAAC 237

Qy         51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
              |||   ::::::      :::       ||||||          :::|||||
Db        238 TATGGACGACTCCGAGGAACATTGAGTGCTTTGCTCTCCTGGTGTCACCTTCATAATAAT 297

Qy         71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIle 90
              :::|||:::|||   |||   :::|||||||||   |||   :::   |||
Db        298 AATTCAACTTTGATCAATAAAATAAACAATCTTCTGGATGCAGTTGGCCAG--------- 348

Qy         91 ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys 110
                           |||:::|||
Db        349 -----------------TGTGAAGAGTAT-------------------------GTG 363

Qy        111 LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu 130
              |||:::   :::   |||||||||      :::      :::|||   |||||::::::
Db        364 CTCAAACATCTGAAATCAATCACTCCAMCGTCCCATGTTGTAGATTTATTGGACTCCATT 423
```

RESULT 15
ABQ56851
ID ABQ56851 standard; cDNA; 677 BP.
XX
AC ABQ56851;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:546.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30732.
XX
PR 02-OCT-2000; 2000US-237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiaglingam A, Lewis ME;
XX
DR WPI; 2002-426115/45.
XX
PT New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell
PT or tissue type, and in antisense therapy -
XX
PS Claim 1; Fig 1; 796pp; English.
XX
CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a

CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridises to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.
XX
SQ Sequence 677 BP; 215 A; 129 C; 166 G; 165 T; 2 other;

Alignment Scores:

Pred. No.:	0.0284	Length:	677
Score:	99.50	Matches:	32
Percent Similarity:	45.83%	Conservative:	23
Best Local Similarity:	26.67%	Mismatches:	46
Query Match:	12.89%	Indels:	19
DB:	24	Gaps:	3

US-09-980-054A-12 (1-148) x ABQ56851 (1-677)

```
Qy          11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
               |||::::::            |||   |||::::::::::|||         |||
Db         129 AACATGCAAAGACGAGTCAAAGAA---GGTTATAGAGATGGAATAGATGCTGGCAAAGCA 185

Qy          31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 50
                  :::|||           ||||||:::   |||   :::       |||
Db         186 GTTACTCTTCAACAGGGCTTCAATCAAGGTTATAAGAAAGGTGCAGAAGTCATTTTAAAC 245

Qy          51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
                 |||   ::::::      :::         ||||||             :::|||||||
Db         246 TATGGACGACTCCGAGGAACATTGAGTGCTTTGCTCTCCTGGTGTCACCTTCATAATAAT 305

Qy          71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIle 90
               :::|||:::|||   |||   :::|||||||||   |||   :::   |||
Db         306 AATTCAACTTTGATCAATAAAATAAACAATCTTCTGGATGCAGTTGGCCAG--------- 356

Qy          91 ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys 110
                              |||:::|||
Db         357 ------------------TGTGAAGAGTAT-------------------------GTG 371

Qy         111 LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu 130
               |||:::   :::   |||||||||     :::      :::|||   ||||||::::::
Db         372 CTCAAACATCTGAAATCAATCACTCCACCGTCCCATGTTGTAGATTTATTGGACTCCATT 431
```

Search completed: February 16, 2004, 14:33:36
Job time : 289 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:24:58 ; Search time 33 Seconds
(without alignments)
939.048 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	764	99.0	148	12	US-10-032-585-7062	Sequence 7062, Ap
2	91.5	11.9	1005	12	US-10-369-493-1061	Sequence 1061, Ap
3	86.5	11.2	876	15	US-10-198-070-56	Sequence 56, Appl
4	86.5	11.2	930	15	US-10-198-070-65	Sequence 65, Appl
5	86	11.1	2285	10	US-09-932-183A-2	Sequence 2, Appli
6	85	11.0	85	12	US-10-369-493-21140	Sequence 21140, A
7	83.5	10.8	688	12	US-10-094-749-1928	Sequence 1928, Ap
8	83.5	10.8	876	15	US-10-198-070-78	Sequence 78, Appl
9	83.5	10.8	879	12	US-10-369-493-1271	Sequence 1271, Ap
10	83.5	10.8	879	12	US-10-369-493-20337	Sequence 20337, A
11	83.5	10.8	880	12	US-10-369-493-21643	Sequence 21643, A
12	83.5	10.8	930	12	US-10-153-244-102	Sequence 102, App
13	83.5	10.8	930	12	US-10-210-152-19	Sequence 19, Appl
14	83.5	10.8	930	15	US-10-005-216-2	Sequence 2, Appli
15	83.5	10.8	930	15	US-10-198-070-73	Sequence 73, Appl
16	83.5	10.8	930	15	US-10-198-070-102	Sequence 102, App
17	83	10.8	1184	9	US-09-815-242-5229	Sequence 5229, Ap
18	83	10.8	1188	9	US-09-815-242-12125	Sequence 12125, A
19	82.5	10.7	435	12	US-10-032-585-7260	Sequence 7260, Ap
20	82.5	10.7	1225	12	US-10-369-493-21875	Sequence 21875, A
21	82	10.6	1164	12	US-10-369-493-6564	Sequence 6564, Ap
22	82	10.6	1963	12	US-10-369-493-5307	Sequence 5307, Ap
23	82	10.6	1963	12	US-10-369-493-5308	Sequence 5308, Ap
24	81	10.5	320	9	US-09-815-242-13785	Sequence 13785, A
25	80.5	10.4	254	12	US-10-032-585-7511	Sequence 7511, Ap
26	80	10.4	235	12	US-10-383-834-5	Sequence 5, Appli
27	80	10.4	457	12	US-10-369-493-5960	Sequence 5960, Ap
28	80	10.4	1156	12	US-10-369-493-43	Sequence 43, Appl
29	80	10.4	1156	12	US-10-289-762-171	Sequence 171, App
30	79.5	10.3	600	12	US-09-882-227-202	Sequence 202, App
31	79	10.2	1173	12	US-10-369-493-21954	Sequence 21954, A
32	78.5	10.2	472	12	US-10-014-099F-63	Sequence 63, Appl
33	78.5	10.2	479	12	US-10-014-099F-67	Sequence 67, Appl
34	78.5	10.2	903	12	US-10-369-493-1048	Sequence 1048, Ap
35	78	10.1	374	12	US-10-032-585-7585	Sequence 7585, Ap
36	77.5	10.0	1143	12	US-10-369-493-11081	Sequence 11081, A
37	77.5	10.0	1173	12	US-10-369-493-5025	Sequence 5025, Ap
38	77	10.0	267	9	US-09-815-242-11549	Sequence 11549, A
39	77	10.0	320	9	US-09-815-242-10422	Sequence 10422, A
40	77	10.0	320	12	US-10-369-493-23638	Sequence 23638, A
41	77	10.0	517	9	US-09-815-242-5387	Sequence 5387, Ap
42	77	10.0	519	9	US-09-815-242-12331	Sequence 12331, A
43	77	10.0	519	9	US-09-815-242-12869	Sequence 12869, A
44	77	10.0	529	12	US-10-301-997-74	Sequence 74, Appl
45	77	10.0	1572	15	US-10-037-182-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-10-032-585-7062
; Sequence 7062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7062
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7062

Query Match 99.0%; Score 764; DB 12; Length 148;
Best Local Similarity 98.6%; Pred. No. 1.5e-69;
Matches 146; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60

Qy         61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120
              ||||||||||||||||||||:|:||||||||||||||||||||||||||||||||||||||
Db         61 WLSHIDQYNNSSSLRNHLNNLENILAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120

Qy        121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
              ||||||||||||||||||||||||||||
Db        121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
```

```
RESULT 2
US-10-369-493-1061
; Sequence 1061, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1061
;   LENGTH: 1005
;   TYPE: PRT
;   ORGANISM: Methanococcus jannaschii
US-10-369-493-1061

  Query Match             11.9%;  Score 91.5;  DB 12;  Length 1005;
  Best Local Similarity   24.7%;  Pred. No. 2.1;
  Matches   43;  Conservative   34;  Mismatches   62;  Indels   35;  Gaps    7;

Qy           3 DID-IDNVLN--------------LEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQR 47
               ||| :| :||                | ||   |  :  :|:  |     | |||     :::
Db         342 DIDNLDTLLNKIKDEIERVETIKDLLEELKNLNEEIEKIEKYKRICEECKEY-----YEK 396

Qy          48 FL-IIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI----------SITNGDKE 96
               :| :     |  |  | :|        |:  ::|:||  : ::           || |  ||
Db         397 YLELEEKAVEYNKLTLEYITLLQEKKSIEKNINDLETRINKLLEETKNIDIESIENSLKE 456

Qy          97 VEDYEK---NIKKARNKL-RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDD 146
               :|: :|   |::| : :| : :  |  |  ::  : : :||| |   : : | |
Db         457 IEEKKKVLENLQKEKIELNKKLGEINSEIKRLKKILDELKEVEGKCPLCKTPID 510


RESULT 3
US-10-198-070-56
; Sequence 56, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT:  GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
;   LENGTH: 876
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-198-070-56

  Query Match             11.2%;  Score 86.5;  DB 15;  Length 876;
  Best Local Similarity   24.2%;  Pred. No. 5.6;
  Matches   38;  Conservative   29;  Mismatches   55;  Indels   35;  Gaps    6;

Qy          19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
                 | ||:     ||| |  : |  |   :   ::| |:|  ::| ||       |:|  | :
Db         723 LKFKKWMSELIQGHKKGFQEDAEMNKRNEEKKFGILGSHEDLSKFSLDRNQLAHNKQSST 782

Qy          71 SSSLRNHLNNLED----------------IMAQISITNGDKEVEDYE----KNIKKARN 109
                 ||   |||:  :                : |||      ||| ::      | ||:  :
Db         783 RSSEDFHLNSFSNPPRQYQKIMKRLIKRYVLQAQI-----DKESDEVNEGELKEIKQDIS 837

Qy         110 KLRVIASITKETWKIDSLDNLVKEVGGTLQVSENPDD 146
                 ||    : :::   : |  |::::|  | :    ::
Db         838 SLR-YELLEEKSQNTEDLAELIRKLGERLSLESKQEE 873


RESULT 4
US-10-198-070-65
; Sequence 65, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT:  GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-198-070-65

  Query Match             11.2%;  Score 86.5;  DB 15;  Length 930;
  Best Local Similarity   24.2%;  Pred. No. 6;
  Matches   38;  Conservative   29;  Mismatches   55;  Indels   35;  Gaps    6;

Qy          19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
                 | ||:     ||| |  : |  |   :   ::| |:|  ::| ||       |:|  | :
Db         777 LKFKKWMSELIQGHKKGFQEDAEMNKRNEEKKFGILGSHEDLSKFSLDRNQLAHNKQSST 836

Qy          71 SSSLRNHLNNLED----------------IMAQISITNGDKEVEDYE----KNIKKARN 109
                 ||   |||:  :                : |||      ||| ::      | ||:  :
Db         837 RSSEDFHLNSFSNPPRQYQKIMKRLIKRYVLQAQI-----DKESDEVNEGELKEIKQDIS 891

Qy         110 KLRVIASITKETWKIDSLDNLVKEVGGTLQVSENPDD 146
                 ||    : :::   : |  |::::|  | :    ::
Db         892 SLR-YELLEEKSQNTEDLAELIRKLGERLSLESKQEE 927


RESULT 5
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
;  APPLICANT: Estell, David A.
;  TITLE OF INVENTION: Proteases From Gram-Positive Organisms
;  FILE REFERENCE: GC394C1-US
;  CURRENT APPLICATION NUMBER: US/09/932,183A
;  CURRENT FILING DATE:  2001-08-17
;  PRIOR APPLICATION NUMBER: US 09/308,375
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: PCT/US98/18828
;  PRIOR FILING DATE: 1998-09-08
;  PRIOR APPLICATION NUMBER: EP9719636.4
;  PRIOR FILING DATE: 1997-09-15
;  NUMBER OF SEQ ID NOS: 3
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2285
;   TYPE: PRT
;   ORGANISM: Bacillius subtilis
US-09-932-183A-2

  Query Match             11.1%;  Score 86;  DB 10;  Length 2285;
  Best Local Similarity   22.4%;  Pred. No. 22;
  Matches   35;  Conservative   34;  Mismatches   61;  Indels   26;  Gaps    6;

Qy          5 DIDNVLNL-EEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGY------IQEL 57
              |:|| ::: ||:: :: :   ||:. : |  | |:|  |   |:    |:      | |
Db       1856 DVDNKISMTEEDEDKVKYYSKQIKLIQQQQKEAKKYIKQLEEQKKAAKGFPDIQEQITEE 1915

Qy         58 MKFWLS-----HIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARN-KL 111
              |: |       ::: ||   |:::   :| | :  |           |::  :| |: :|
Db       1916 MQNWKDKQKDFNLELYNTKKSIKDIYKSLADEVVSI----------YKEMYEKMRDIEL 1964

Qy        112 RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDDM 147
                  |::   || :|    |     :: |  | :
Db       1965 EAHQKATQDL--IDEIDKTDDEAKFQKELKERQDSI 1998


RESULT 6
US-10-369-493-21140
; Sequence 21140, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21140
;   LENGTH: 85
;   TYPE: PRT
;   ORGANISM: Xenorhabdus nematophilus
US-10-369-493-21140

  Query Match             11.0%;  Score 85;  DB 12;  Length 85;
  Best Local Similarity   32.9%;  Pred. No. 0.37;
  Matches   28;  Conservative   13;  Mismatches   34;  Indels   10;  Gaps    4;

Qy         65 IDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKI 124
              : |  ||  ||  | :| | : ::  ::|||   : ||   ||   |:  | :|
Db          1 MSQKKNSEDLRTELQSLADTLEEVLNSSGDKSKAELEKLRSKAEKTLK-DARVTLS---- 55

Qy        125 DSLDNLV---KEVGGTLQ--VSENP 144
              |: | ||   ||: |     | :||
Db         56 DASDKLVDQTKEMAGRADNYVRDNP 80


RESULT 7
US-10-094-749-1928
; Sequence 1928, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT:  SUGIYAMA, TOMOYASU
;  APPLICANT:  OTSUKI, TETSUJI
;  APPLICANT:  WAKAMATSU, AI
;  APPLICANT:  SATO, HIROYUKI
;  APPLICANT:  ISHII, SHIZUKO
;  APPLICANT:  YAMAMOTO, JUN-ICHI
;  APPLICANT:  ISONO, YUUKO
;  APPLICANT:  HIO, YURI
;  APPLICANT:  OTSUKA, KAORU
;  APPLICANT:  NAGAI, KEIICHI
;  APPLICANT:  IRIE, RYOTARO
;  APPLICANT:  TAMECHIKA, ICHIRO
;  APPLICANT:  SEKI, NAOHIKO
;  APPLICANT:  YOSHIKAWA, TSUTOMU
;  APPLICANT:  OTSUKA, MOTOYUKI
;  APPLICANT:  NAGAHARI, KENJI
;  APPLICANT:  MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1928
;   LENGTH: 688
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-749-1928

  Query Match             10.8%;  Score 83.5;  DB 12;  Length 688;
  Best Local Similarity   22.2%;  Pred. No. 8.1;
  Matches   35;  Conservative   32;  Mismatches   48;  Indels   43;  Gaps    7;

Qy          5 DIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              | : | || :|| :|  :|       | | :|      :||  |        :  | | :  |
Db        126 DDETVRNL-QEQLQLANQE------KTQAVE----LWQTVSQE------LDRLHKLYQEH 168

Qy         65 IDQYN----NSSSLRNHLNNLEDIMAQISITNGDKEV----------------EDYEKNI 104
              : :        |   ::  | : : :  |: :|| : ||                |   | :
Db        169 MTEAQIHVFESQKQKDQLFDFQQLTKQLHVTNENMEVTNQQFLKTVTEQSVIIEQLRKKL 228

Qy        105 KKARNKLRVIASITKETWKIDSLDNLVKEVGGTLQVSE 142
              ::|: :|||  :      |:: | |: ::: | ::  |
Db        229 RQAKLELRVAVA------KVEELTNVTEDLQGQMKKKE 260


RESULT 8
US-10-198-070-78
; Sequence 78, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT:  GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
;   LENGTH: 876
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-198-070-78

  Query Match             10.8%;  Score 83.5;  DB 15;  Length 876;
  Best Local Similarity   25.3%;  Pred. No. 11;
```

Matches 37; Conservative 29; Mismatches 49; Indels 31; Gaps 6;

```
Qy          19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
               | ||:     |||  |  : |   |    :   ::| | |  ::| ||        |:|  | :
Db         723 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 782

Qy          71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
                ||   |||          | :|  ::   |:| :|:   :  : | | ||: :::  :
Db         783 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 828

Qy         131 VKEVGGTL---------QVSENPDDM 147
               :||:   :         : |:| :|:
Db         829 LKEIKQDISSLRYELLEEKSQNTEDL 854
```

RESULT 9
US-10-369-493-1271
; Sequence 1271, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1271
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1271

Query Match 10.8%; Score 83.5; DB 12; Length 879;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 37; Conservative 40; Mismatches 61; Indels 49; Gaps 6;

```
Qy           1 MSDIDIDNVLNLEEEQYELGFKE-----GQIQGTKDQYLEGKEYGYQT------------ 43
               :| |:::: |  ::|:|||   |      |::  | :  |  :|  ::
Db         512 LSKINLED-LKRDKEEYELLKSESNKLKGEVESLKKEVNELNDYKNESTKLEIEIDKAKK 570

Qy          44 ------------GFQRF-LIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI-- 88
                         ||:    : | |:|| ||   :|:  |    ||: | :|:|   ::
Db         571 ELSEIEDRLLRLGFKTIDELSGRIRELEKFHNKYIEAKNAEKELRDILESLKDEREELDK 630

Qy          89 -------------SITNGDKEVE---DYEKNIKKARNKLRVIASITKETWKIDSLDNLVK 132
                          :|:   |::   | :|  :|    :::   |      |:: |:
Db         631 AFEELAKIETDIEKVTSQLNELQRKFDQKKYEEKREKMMKLSMEIKGLETKLEELERRRD 690

Qy         133 EVGGTLQ 139
               |:  |::
Db         691 EIKSTIE 697
```

RESULT 10
US-10-369-493-20337
; Sequence 20337, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20337
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20337

Query Match 10.8%; Score 83.5; DB 12; Length 879;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 37; Conservative 40; Mismatches 61; Indels 49; Gaps 6;

```
Qy           1 MSDIDIDNVLNLEEEQYELGFKE-----GQIQGTKDQYLEGKEYGYQT------------ 43
               :| |:::: |  ::|:|||   |      |::  | :  |  :|  ::
Db         512 LSKINLED-LKRDKEEYELLKSESNKLKGEVESLKKEVNELNDYKNESTKLEIEIDKAKK 570

Qy          44 ------------GFQRF-LIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI-- 88
                         ||:    : | |:|| ||   :|:  |    ||: | :|:|   ::
Db         571 ELSEIEDRLLRLGFKTIDELSGRIRELEKFHNKYIEAKNAEKELRDILESLKDEREELDK 630

Qy          89 -------------SITNGDKEVE---DYEKNIKKARNKLRVIASITKETWKIDSLDNLVK 132
                          :|:   |::   | :|  :|    :::   |      |:: |:
Db         631 AFEELAKIETDIEKVTSQLNELQRKFDQKKYEEKREKMMKLSMEIKGLETKLEELERRRD 690

Qy         133 EVGGTLQ 139
               |:  |::
Db         691 EIKSTIE 697
```

RESULT 11
US-10-369-493-21643
; Sequence 21643, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21643
; LENGTH: 880
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21643

Query Match 10.8%; Score 83.5; DB 12; Length 880;
Best Local Similarity 20.1%; Pred. No. 11;
Matches 38; Conservative 35; Mismatches 49; Indels 67; Gaps 8;

```
Qy          11 NLEEEQYELGFKEGQIQGTKDQY--LEGKEYGYQTGFQRF--------LI---------- 50
               |||    ||  || : :|  :::  |:|:  | :   :|         ||
Db         517 NLE----ELEQKEREFEGLNEEFNKLKGELLGLERDLKRIKALEGRRKLIEEKVRKAKEE 572

Qy          51 ----------IGY---------IQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQIS-- 89
                        :|:         |||| :|   :::   :  | ||   | ||    ::
Db         573 LENLHRQLRELGFESVEELNLRIQELEEFHDKYVEAKKSESELRELKNKLEKEKTELDQA 632

Qy          90 ---ITNGDKEVEDYEKNIKKARNKL---------------RVIASITKETWKIDSLDNL 130
                  : : : |:|: |  :|   :|                 | ::|:|    ::: |
Db         633 FEMLADVENEIEEKEAKLKDLESKFNEEEYEEKRERLVKLEREVSSLTA---RLEELKKS 689

Qy         131 VKEVGGTLQ 139
```

```
              |:::  ||:
Db        690 VEQIKATLR 698


RESULT 12
US-10-153-244-102
; Sequence 102, Application US/10153244
; Publication No. US20030144191A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL TRP CHANNEL FAMILY MEMBER, TRP-PL
;  TITLE OF INVENTION:  SPLICE VARIANTS THEREOF
;  FILE REFERENCE: D0144 NP
;  CURRENT APPLICATION NUMBER: US/10/153,244
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,599
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/362,944
;  PRIOR FILING DATE: 2002-03-08
;  NUMBER OF SEQ ID NOS: 335
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 102
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-153-244-102

  Query Match             10.8%;  Score 83.5;  DB 12;  Length 930;
  Best Local Similarity   25.3%;  Pred. No. 12;
  Matches   37;  Conservative   29;  Mismatches   49;  Indels   31;  Gaps    6;

Qy         19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
              | ||:     ||| |  : |  |   :   ::| | |  ::| ||      |:|  | :
Db        777 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 836

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              ||   |||         | :|  ::   |:| :|:   :  : | | ||: :::  :
Db        837 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 882

Qy        131 VKEVGGTL---------QVSENPDDM 147
              :||:   :         : |:| :|:
Db        883 LKEIKQDISSLRYELLEEKSQNSEDL 908


RESULT 13
US-10-210-152-19
; Sequence 19, Application US/10210152
; Publication No. US20030162189A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL TRP CHANNEL FAMILY MEMBER, LTRPC3
;  TITLE OF INVENTION:  SPLICE VARIANTS THEREOF
;  FILE REFERENCE: D0171 NP
;  CURRENT APPLICATION NUMBER: US/10/210,152
;  CURRENT FILING DATE:  2002-08-01
;  PRIOR APPLICATION NUMBER: US 60/309,544
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 320
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-210-152-19

  Query Match             10.8%;  Score 83.5;  DB 12;  Length 930;
  Best Local Similarity   25.3%;  Pred. No. 12;
  Matches   37;  Conservative   29;  Mismatches   49;  Indels   31;  Gaps    6;

Qy         19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
              | ||:     ||| |  : |  |   :   ::| | |  ::| ||      |:|  | :
Db        777 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 836

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              ||   |||         | :|  ::   |:| :|:   :  : | | ||: :::  :
Db        837 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 882

Qy        131 VKEVGGTL---------QVSENPDDM 147
              :||:   :         : |:| :|:
Db        883 LKEIKQDISSLRYELLEEKSQNSEDL 908


RESULT 14
US-10-005-216-2
; Sequence 2, Application US/10005216
; Publication No. US20030079243A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Keith D.
;  TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CALCIUM ION
;  TITLE OF INVENTION:  CHANNEL (Trp6) GENE DISRUPTIONS
;  FILE REFERENCE: R-881
;  CURRENT APPLICATION NUMBER: US/10/005,216
;  CURRENT FILING DATE:  2001-12-04
;  PRIOR APPLICATION NUMBER: US 60/280,373
;  PRIOR FILING DATE: 2001-03-29
;  PRIOR APPLICATION NUMBER: US 60/255,227
;  PRIOR FILING DATE: 2000-12-11
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-005-216-2

  Query Match             10.8%;  Score 83.5;  DB 15;  Length 930;
  Best Local Similarity   25.3%;  Pred. No. 12;
  Matches   37;  Conservative   29;  Mismatches   49;  Indels   31;  Gaps    6;

Qy         19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
              | ||:     ||| |  : |  |   :   ::| | |  ::| ||      |:|  | :
Db        777 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 836

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              ||   |||         | :|  ::   |:| :|:   :  : | | ||: :::  :
Db        837 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 882

Qy        131 VKEVGGTL---------QVSENPDDM 147
              :||:   :         : |:| :|:
Db        883 LKEIKQDISSLRYELLEEKSQNSEDL 908


RESULT 15
US-10-198-070-73
; Sequence 73, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
;  APPLICANT: AVERBACK, PAUL
;  APPLICANT:  GEMMELL, JACK
;  TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
;  TITLE OF INVENTION:  CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 59003.000008
;  CURRENT APPLICATION NUMBER: US/10/198,070
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: 60/306,161
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/306,150
;  PRIOR FILING DATE: 2001-07-19
;  PRIOR APPLICATION NUMBER: 60/331,477
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 125
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
;   LENGTH: 930
;   TYPE: PRT
```

; ORGANISM: Mus musculus
US-10-198-070-73

Query Match 10.8%; Score 83.5; DB 15; Length 930;
Best Local Similarity 25.3%; Pred. No. 12;
Matches 37; Conservative 29; Mismatches 49; Indels 31; Gaps 6;

```
Qy          19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
               | ||:     ||| |  : |  |    :   ::| | |  ::| ||       |:|   | :
Db         777 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 836

Qy          71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
                ||   |||          | :|  ::   |:| :|:   :  : | | ||: :::  :
Db         837 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 882

Qy         131 VKEVGGTL---------QVSENPDDM 147
               :||:   :         : |:| :|:
Db         883 LKEIKQDISSLRYELLEEKSQNTEDL 908
```

Search completed: February 16, 2004, 14:28:33
Job time : 34 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:12:37 ; Search time 28 Seconds
(without alignments)
223.643 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	11.1	2285	4	US-09-308-375-2	Sequence 2, Appli
2	82	10.6	756	4	US-09-134-001C-3553	Sequence 3553, Ap
3	81	10.5	521	4	US-09-508-370A-3	Sequence 3, Appli
4	80.5	10.4	275	2	US-08-392-625-17	Sequence 17, Appl
5	80.5	10.4	275	2	US-08-466-961A-17	Sequence 17, Appl
6	80	10.4	1156	4	US-09-198-452A-171	Sequence 171, App
7	78	10.1	506	4	US-09-252-991A-23066	Sequence 23066, A
8	78	10.1	1104	3	US-08-923-992A-4	Sequence 4, Appli
9	77.5	10.0	496	3	US-08-867-611-10	Sequence 10, Appl
10	77.5	10.0	496	5	PCT-US92-06965A-15	Sequence 15, Appl
11	77.5	10.0	498	1	US-08-500-222-2	Sequence 2, Appli
12	77.5	10.0	498	1	US-08-500-125-2	Sequence 2, Appli
13	77.5	10.0	498	2	US-07-779-704B-2	Sequence 2, Appli
14	77	10.0	425	3	US-08-867-611-22	Sequence 22, Appl
15	77	10.0	425	5	PCT-US92-06965A-27	Sequence 27, Appl
16	77	10.0	497	4	US-09-328-352-7286	Sequence 7286, Ap
17	77	10.0	529	4	US-08-887-534A-74	Sequence 74, Appl
18	77	10.0	529	4	US-09-527-431-74	Sequence 74, Appl
19	77	10.0	617	3	US-08-867-611-30	Sequence 30, Appl
20	77	10.0	617	5	PCT-US92-06965A-35	Sequence 35, Appl
21	76.5	9.9	496	3	US-08-926-842B-12	Sequence 12, Appl
22	76.5	9.9	810	4	US-09-540-824-25	Sequence 25, Appl
23	76	9.8	396	1	US-08-430-024-2	Sequence 2, Appli
24	76	9.8	396	1	US-08-782-009-2	Sequence 2, Appli
25	76	9.8	396	3	US-09-017-302-2	Sequence 2, Appli
26	76	9.8	1164	3	US-08-923-992A-2	Sequence 2, Appli
27	75.5	9.8	258	4	US-09-107-532A-4489	Sequence 4489, Ap
28	75.5	9.8	330	4	US-09-134-001C-3811	Sequence 3811, Ap
29	75.5	9.8	1211	4	US-09-134-001C-4820	Sequence 4820, Ap
30	75	9.7	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
31	74.5	9.7	255	4	US-08-858-207A-465	Sequence 465, App
32	74.5	9.7	263	1	US-08-160-670A-32	Sequence 32, Appl
33	74.5	9.7	264	1	US-08-160-670A-28	Sequence 28, Appl
34	74.5	9.7	488	1	US-08-190-802A-60	Sequence 60, Appl
35	74.5	9.7	488	3	US-08-477-346-60	Sequence 60, Appl
36	74.5	9.7	488	4	US-08-473-089-60	Sequence 60, Appl
37	74.5	9.7	488	4	US-08-487-072A-60	Sequence 60, Appl
38	74.5	9.7	527	1	US-08-160-670A-8	Sequence 8, Appli
39	74.5	9.7	528	1	US-08-160-670A-5	Sequence 5, Appli
40	74	9.6	466	2	US-08-912-129A-55	Sequence 55, Appl
41	73.5	9.5	244	2	US-08-553-633A-5	Sequence 5, Appli
42	73.5	9.5	580	4	US-09-256-000-21	Sequence 21, Appl
43	73.5	9.5	593	2	US-08-987-466-1	Sequence 1, Appli
44	73.5	9.5	593	3	US-09-240-359-1	Sequence 1, Appli
45	73.5	9.5	983	4	US-09-134-001C-3814	Sequence 3814, Ap

ALIGNMENTS

RESULT 1
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillius subtilis
US-09-308-375-2

Query Match 11.1%; Score 86; DB 4; Length 2285;
Best Local Similarity 22.4%; Pred. No. 3.9;
Matches 35; Conservative 34; Mismatches 61; Indels 26; Gaps 6;

```
Qy          5 DIDNVLNL-EEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGY------IQEL 57
              |:|| ::: ||:: :: :   ||:  : |  | |:| |  |:   |:      | ||
Db       1856 DVDNKISMTEEDEDKVKYYSKQIKLIQQQQKEAKKYIKQLEEQKKAAKGFPDIQEQITEE 1915

Qy         58 MKFWLS-----HIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARN-KL 111
              |: |         ::: ||   |:::   :| | :  |           |::  :| |: :|
Db       1916 MQNWKDKQKDFNLELYNTKKSIKDIYKSLADEVVSI----------YKEMYEKMRDIEL 1964

Qy        112 RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDDM 147
                 |::   || :|    |     :: |  | :
Db       1965 EAHQKATQDL--IDEIDKTDDEAKFQKELKERQDSI 1998
```

RESULT 2
US-09-134-001C-3553
; Sequence 3553, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08

```
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3553
;   LENGTH: 756
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3553

  Query Match             10.6%;  Score 82;  DB 4;  Length 756;
  Best Local Similarity   23.0%;  Pred. No. 2.3;
  Matches   35;  Conservative   31;  Mismatches   58;  Indels   28;  Gaps    6;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQ-----TGFQRFLIIGYIQELMKFW 61
              ::|: | || |      | :  :|    |::||:      | |:       :| |   :
Db        217 EDVIRLREELSE------QYRSLQDLKELGQKYGFDISRPATNFKE-----AVQWLYLAY 265

Qy         62 LSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIA-SITKE 120
              |: | : | ::     :   || |:  : |||   :: :: |     ||::  : | |
Db        266 LAAIKEQNGAAMSLGRTSTFLDIYAERDLQNGDITEQEVQEIIDHFIMKLRIVKFARTPE 325

Qy        121 ---------TWKIDSLDNLVKEVGGTLQVSEN 143
                      ||  :|:  :   : |   |::|
Db        326 YNELFSGDPTWVTESIGGV--GIDGRPMVTKN 355


RESULT 3
US-09-508-370A-3
; Sequence 3, Application US/09508370A
; Patent No. 6492131
; GENERAL INFORMATION:
;  APPLICANT:      Dieter Soll
;  APPLICANT:  Michael Ibba
;  TITLE OF INVENTION:      Class I-type Lysyl-tRNA Synthetase
;  FILE REFERENCE:      OCR-896
;  CURRENT APPLICATION NUMBER: US/09/508,370A
;  CURRENT FILING DATE:  2000-03-10
;  PRIOR APPLICATION NUMBER:      PCT/US98/18968
;  PRIOR FILING DATE:      1998-09-09
;  NUMBER OF SEQ ID NOS:      16
;  SOFTWARE:      MS Dos
; SEQ ID NO 3
;   LENGTH: 521
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: lysyl t-RNA synthetase
;   OTHER INFORMATION: construct expressed in Example 3
US-09-508-370A-3

  Query Match             10.5%;  Score 81;  DB 4;  Length 521;
  Best Local Similarity   21.5%;  Pred. No. 1.7;
  Matches   32;  Conservative   22;  Mismatches   49;  Indels   46;  Gaps    4;

Qy         14 EEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSS 73
              :|: :  ||  :|      |:  |   || ||:   :|  | |             || :
Db        345 KEEKKRAFK--RIYELSQPYMPSKRIPYQVGFRHLSVISQIPE-----------NNINK 390

Qy         74 LRNHLNNLEDIMA---------------------QISITNGDKEVEDYEKNIKKARNKL 111
              : |:|  |:::                      : |: :    :|  |:| ||| |:|
Db        391 ILNYLKNVQEDQKDKLINKINCAINWIRDFAPEDFKFSLRSKFDNMEILEENSKKAINEL 450

Qy        112 ----------RVIASITKETWKIDSLDNL 130
                       |  | :||   :|:
Db        451 LDFLKKNFEVATEQDIQNEIYKISRENNI 479


RESULT 4
US-08-392-625-17
; Sequence 17, Application US/08392625
; Patent No. 5837485
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5837485bert
;    APPLICANT:  Augustin, Johannes
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process For The Preparation
;    TITLE OF INVENTION:  Of Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/392,625
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-392-625-17

  Query Match             10.4%;  Score 80.5;  DB 2;  Length 275;
  Best Local Similarity   27.5%;  Pred. No. 0.8;
  Matches   36;  Conservative   24;  Mismatches   44;  Indels   27;  Gaps    8;

Qy         21 FKEGQIQGTKDQYLEGKEYGYQTGFQRFL------------IIGYIQ-ELMKFWLSHIDQ 67
              :|  |    || |: |    |:|     :             || ||  : :   :|:| |
Db         13 YKSFQSDLKKDLYINGT---YETNLGALINLLTSKGCGSEGIINYIYGKHLDLSISNIIQ 69

Qy         68 YNNSSSLRNHLNNLEDIMAQI--SITNGDKEVEDYEKNIKKARNKLRVIASITKETW--K 123
                 |: :    :|:|:|  | |:  : |      ||: |||||  |:   :   :  |||
Db         70 KNDKTF---NLDNIE--MKQVFNNYTTKITLVEEPEKNIKLIRDSSGIARQV--ETWYST 122

Qy        124 IDSLDNLVKEV 134
                |  : |:::|:
Db        123 ISVIPNIIQEL 133


RESULT 5
US-08-466-961A-17
; Sequence 17, Application US/08466961A
; Patent No. 5843709
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5843709bert
;    APPLICANT:  Augustin, Johannes
```

```
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process for the Preparation of
;    TITLE OF INVENTION:  Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,961A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/392,625
;      FILING DATE:  22-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/784,234
;      FILING DATE:  31-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-466-961A-17

  Query Match             10.4%;  Score 80.5;  DB 2;  Length 275;
  Best Local Similarity   27.5%;  Pred. No. 0.8;
  Matches   36;  Conservative   24;  Mismatches   44;  Indels   27;  Gaps    8;

Qy         21 FKEGQIQGTKDQYLEGKEYGYQTGFQRFL-----------IIGYIQ-ELMKFWLSHIDQ 67
              :| |    || |: |    |:|     :           || ||  : :   :|:| |
Db         13 YKSFQSDLKKDLYINGT---YETNLGALINLLTSKGCGSEGIINYIYGKHLDLSISNIIQ 69

Qy         68 YNNSSSLRNHLNNLEDIMAQI--SITNGDKEVEDYEKNIKKARNKLRVIASITKETW--K 123
               |: :    :|:|:|  | |:  : |     ||: |||||  |:   :   :  |||
Db         70 KNDKTF---NLDNIE--MKQVFNNYTTKITLVEEPEKNIKLIRDSSGIARQV--ETWYST 122

Qy        124 IDSLDNLVKEV 134
               |  : |:::|:
Db        123 ISVIPNIIQEL 133


RESULT 6
US-09-198-452A-171
; Sequence 171, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 171
;   LENGTH: 1156
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-198-452A-171

  Query Match             10.4%;  Score 80;  DB 4;  Length 1156;
  Best Local Similarity   28.3%;  Pred. No. 6.8;
  Matches   39;  Conservative   16;  Mismatches   47;  Indels   36;  Gaps    6;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQY--LEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              | || |:          | |:|| ||     |||                ||  |  ||
Db        155 DEVLALKSTM------EAQLQGFKDLVGTWEGK----------------YQEFKKNKLSK 192

Qy         65 IDQYNNSSSLRNHLNNLEDIMAQIS----ITNGDKEVEDYEKNIKKARNKLRVIASITKE 120
              :  |: : |  | || || : |: |    : : |:  || :| |::    |     :| |
Db        193 VLVYDFTKSYSNLLNRLEVLHAESSTDDLVLHVDRMSEDLKKTIEEIDGNL---FQVTPE 249

Qy        121 TWKIDSLDNLVKEVGGTL 138
                    | | :|  | :
Db        250 -----ELSLLAREYQGLM 262


RESULT 7
US-09-252-991A-23066
; Sequence 23066, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 23066
;   LENGTH: 506
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23066

  Query Match             10.1%;  Score 78;  DB 4;  Length 506;
  Best Local Similarity   22.1%;  Pred. No. 3.5;
  Matches   29;  Conservative   22;  Mismatches   34;  Indels   46;  Gaps    5;

Qy         29 TKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLED----- 83
              :::| :|   :|   |   ||  | :|||              | || ||| |:|
Db        148 SREQVVEQLHFG--GGTPTFLSPGQLRELM------------SQLRTHLNLLDDDSGDY 192

Qy         84 --------------IMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLD 128
                           :: ::          |:|::  ::|| |:::      :||
Db        193 GIEIDPREADWSTMGLLRELGFNRVSLGVQDFDMEVQKAVNRMQT----PEET------- 241

Qy        129 NLVKEVGGTLQ 139
                : |   |||
Db        242 RTIVEAARTLQ 252


RESULT 8
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Blake, Milan S.
;    TITLE OF INVENTION:  No. 6280738-IgA Fc Binding Forms of the Group B
;    TITLE OF INVENTION:  Streptococcal Beta Antigens
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,992A
;      FILING DATE:  05-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,707
;      FILING DATE:  06-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  1438.0140001/RWE
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1104 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-923-992A-4

  Query Match             10.1%;  Score 78;  DB 3;  Length 1104;
  Best Local Similarity   24.0%;  Pred. No. 10;
  Matches   36;  Conservative   26;  Mismatches   64;  Indels   24;  Gaps    5;

Qy          5 DIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQEL------- 57
              |:|| :   |:: |   |  :   : |  || :    :|   | | |   :|:
Db         90 DVDNTILSHEQKNEFKTKIDETNDS-DALLELENQFNET--NRLLHIKQHEEVEKHNKPN 146

Qy         58 ----MKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKE-----VEDYEKNIKKAR 108
                  :|   : :|  |    | :  : :| :  |  ||| ||:     :|| |  ::|
Db        147 QQKTLKQSDTKVDLSNIDKELNHQKSQVEAMAEQAGITNEDKDSMLKKIEDIRKQAQQAD 206

Qy        109 NKLRVIASITKETWKIDS-----LDNLVKE 133
                  |      : :|  |: |     ||  ::|
Db        207 KKEDAEVKVREELGKLFSSTKAGLDQQIQE 236


RESULT 9
US-08-867-611-10
; Sequence 10, Application US/08867611
; Patent No. 6172189
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, SUSHIL G
;    APPLICANT:  DESAI, SURESH M
;    APPLICANT:  CASEY, JAMES M
;    APPLICANT:  DAILEY, STEPHEN H
;    APPLICANT:  DAWSON, GEORGE J
;    APPLICANT:  GUTIERREZ, ROBIN A
;    APPLICANT:  LESNIEWSKI, RICHARD R
;    APPLICANT:  STEWART, JAMES L
;    APPLICANT:  RUPPRECHT, KEVIN R
;    TITLE OF INVENTION:  HEPATITIS C ASSAY UTILIZING RECOMBINANT
;    TITLE OF INVENTION:  ANTIGENS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD, CHAD377/AP6D2
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,611
;      FILING DATE:  02-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,757
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/179,896
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/572,822
;      FILING DATE:  24-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/614,069
;      FILING DATE:  07-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,561
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,565
;      FILING DATE:  21-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/748,566
;      FILING DATE:  21-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834.US.P6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  496 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-867-611-10

  Query Match             10.0%;  Score 77.5;  DB 3;  Length 496;
  Best Local Similarity   20.4%;  Pred. No. 3.9;
  Matches   37;  Conservative   22;  Mismatches   67;  Indels   55;  Gaps    6;

Qy         12 LEEEQYELGFKEGQIQGTKDQYLEGKE-----------YGYQTGFQRFLI--------- 50
              |: | | | |     |   :|:: || | |           |||: || |  :
Db        146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy         51 IGYIQELMKFWLS---HI--------------DQYNNSSSLRNHLN------------ 79
              |  :::|   |     |:               |   :::|| :|
Db        206 IEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTPEDLDPSTNSLDSHYQDVLKEVKAAASKV 265

Qy         80 --NLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNLVKEVGGT 137
               ||  :    |:|        :   |  |   |    :      || | |::  |  :  |
Db        266 KANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVTHINSVWK-DLLEDNVTPIDTT 324

Qy        138 L 138
               :
Db        325 I 325
```

```
RESULT 10
PCT-US92-06965A-15
; Sequence 15, Application PC/TUS9206965A
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, S.
;    APPLICANT:  DESAI, S.
;    APPLICANT:  DAILEY, S.
;    TITLE OF INVENTION:  HCV SYNTHETIC PEPTIDE FROM NS1 REGION
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  U.S.
;      ZIP:  60065-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06965A
;      FILING DATE:  19920821
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834PC.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  496 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-06965A-15

  Query Match             10.0%;  Score 77.5;  DB 5;  Length 496;
  Best Local Similarity   20.4%;  Pred. No. 3.9;
  Matches   37;  Conservative   22;  Mismatches   67;  Indels   55;  Gaps    6;

Qy          12 LEEEQYELGFKEGQIQGTKDQYLEGKE----------YGYQTGFQRFLI--------- 50
               |: | | | |     |   :|:: || |            |||: || |  :
Db         146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy          51 IGYIQELMKFWLS---HI--------------DQYNNSSSLRNHLN------------ 79
               |  :::|   |     |:               |   :::|| :|
Db         206 IEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTPEDLDPSTNSLDSHYQDVLKEVKAAASKV 265

Qy          80 --NLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNLVKEVGGT 137
                 ||  :   |:|     :   |  |   |   :     || | |:: |  :  |
Db         266 KANLLSVEEACSLTPPHSAKSKFGYGAKDVRCHARKAVTHINSVWK-DLLEDNVTPIDTT 324

Qy         138 L 138
               :
Db         325 I 325


RESULT 11
US-08-500-222-2
; Sequence 2, Application US/08500222
; Patent No. 5643733
;  GENERAL INFORMATION:
;    APPLICANT:  Robinson, John M
;    APPLICANT:  Pilot-Matias, Tami J
;    APPLICANT:  Hunt, Jeffrey C
;    TITLE OF INVENTION:  Borrelia burgdorferi antigens
;    TITLE OF INVENTION:  and uses thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  One Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/500,222
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/779,704
;      FILING DATE:  21-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wong, Wean Khing
;      REGISTRATION NUMBER:  33561
;      REFERENCE/DOCKET NUMBER:  5051.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  498 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Borrelia burgdorferi
;      STRAIN:  B31
US-08-500-222-2

  Query Match             10.0%;  Score 77.5;  DB 1;  Length 498;
  Best Local Similarity   19.8%;  Pred. No. 3.9;
  Matches   34;  Conservative   27;  Mismatches   46;  Indels   65;  Gaps    8;

Qy          12 LEEEQYELGFKEGQIQGTKDQYLEGKE------------YGYQTGFQRFLI--------- 50
               |: | | | |     |   :|:: || |            |||: || |  :
Db         146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy          51 IGYIQELMKFWLS---HI--------------------------DQYNNSSSL------RN 76
               |  :::|   |     |:                         |  | | |  ::
Db         206 IEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTPENPSTGLMKISDPRNTSKAINFIQTTEG 265

Qy          77 HLNNLEDI---MAQISITNGDKEVEDYEKNIKKARNKLRV-IASITKETWKI 124
               :|| :| :   | :::: :|:    | :      |  ::: |  :| |  :|
Db         266 NLNEVEKVLVRMKELAVQSGNGTYSDAD------RGSIQIEIEQLTDEINRI 311


RESULT 12
US-08-500-125-2
; Sequence 2, Application US/08500125
; Patent No. 5643751
;  GENERAL INFORMATION:
;    APPLICANT:  Robinson, John M
;    APPLICANT:  Pilot-Matias, Tami J
;    APPLICANT:  Hunt, Jeffrey C
;    TITLE OF INVENTION:  Borrelia burgdorferi antigens
;    TITLE OF INVENTION:  and uses thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  One Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
```

```
;       ZIP:  60064
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/500,125
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/779,704
;       FILING DATE:  21-OCT-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wong, Wean Khing
;       REGISTRATION NUMBER:  33561
;       REFERENCE/DOCKET NUMBER:  5051.US.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  708-937-9396
;       TELEFAX:  708-937-9556
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  498 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Borrelia burgdorferi
;       STRAIN:  B31
US-08-500-125-2

  Query Match             10.0%;  Score 77.5;  DB 1;  Length 498;
  Best Local Similarity   19.8%;  Pred. No. 3.9;
  Matches   34;  Conservative   27;  Mismatches   46;  Indels   65;  Gaps    8;

Qy          12 LEEEQYELGFKEGQIQGTKDQYLEGKE------------YGYQTGFQRFLI--------- 50
               |: | | | |     |   :|:: || |            |||: || |  :
Db         146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy          51 IGYIQELMKFWLS---HI-----------------------DQYNNSSSL------RN 76
               |  :::|   |     |:                      |  | | | ::
Db         206 IEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTPENPSTGLMKISDPRNTSKAINFIQTTEG 265

Qy          77 HLNNLEDI---MAQISITNGDKEVEDYEKNIKKARNKLRV-IASITKETWKI 124
               :|| :| :   | :::: :|:    | :      |  ::: |  :| |  :|
Db         266 NLNEVEKVLVRMKELAVQSGNGTYSDAD------RGSIQIEIEQLTDEINRI 311


RESULT 13
US-07-779-704B-2
; Sequence 2, Application US/07779704B
; Patent No. 5965702
;  GENERAL INFORMATION:
;    APPLICANT:  Robinson, John M
;    APPLICANT:  Pilot-Matias, Tami J
;    APPLICANT:  Hunt, Jeffrey C
;    TITLE OF INVENTION:  Borrelia burgdorferi antigens
;    TITLE OF INVENTION:  and uses thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  One Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/779,704B
;      FILING DATE:  19911021
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wong, Wean Khing
;      REGISTRATION NUMBER:  33561
;      REFERENCE/DOCKET NUMBER:  5051.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  498 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Borrelia burgdorferi
;      STRAIN:  B31
US-07-779-704B-2

  Query Match             10.0%;  Score 77.5;  DB 2;  Length 498;
  Best Local Similarity   19.8%;  Pred. No. 3.9;
  Matches   34;  Conservative   27;  Mismatches   46;  Indels   65;  Gaps    8;

Qy          12 LEEEQYELGFKEGQIQGTKDQYLEGKE------------YGYQTGFQRFLI--------- 50
               |: | | | |     |   :|:: || |            |||: || |  :
Db         146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy          51 IGYIQELMKFWLS---HI-----------------------DQYNNSSSL------RN 76
               |  :::|   |     |:                      |  | | | ::
Db         206 IEMLEQLRVLWYGEKIHVAVAQEVPGTGVDTPENPSTGLMKISDPRNTSKAINFIQTTEG 265

Qy          77 HLNNLEDI---MAQISITNGDKEVEDYEKNIKKARNKLRV-IASITKETWKI 124
               :|| :| :   | :::: :|:    | :      |  ::: |  :| |  :|
Db         266 NLNEVEKVLVRMKELAVQSGNGTYSDAD------RGSIQIEIEQLTDEINRI 311


RESULT 14
US-08-867-611-22
; Sequence 22, Application US/08867611
; Patent No. 6172189
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, SUSHIL G
;    APPLICANT:  DESAI, SURESH M
;    APPLICANT:  CASEY, JAMES M
;    APPLICANT:  DAILEY, STEPHEN H
;    APPLICANT:  DAWSON, GEORGE J
;    APPLICANT:  GUTIERREZ, ROBIN A
;    APPLICANT:  LESNIEWSKI, RICHARD R
;    APPLICANT:  STEWART, JAMES L
;    APPLICANT:  RUPPRECHT, KEVIN R
;    TITLE OF INVENTION:  HEPATITIS C ASSAY UTILIZING RECOMBINANT
;    TITLE OF INVENTION:  ANTIGENS
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD, CHAD377/AP6D2
;      CITY:  ABBOTT PARK
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/867,611
;       FILING DATE:  02-JUN-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/646,757
;       FILING DATE:
;       APPLICATION NUMBER:  US/08/179,896
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/572,822
;       FILING DATE:  24-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/614,069
;       FILING DATE:  07-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/748,561
;       FILING DATE:  21-AUG-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/748,565
;       FILING DATE:  21-AUG-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/748,566
;       FILING DATE:  21-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  POREMBSKI, PRISCILLA E
;       REGISTRATION NUMBER:  33,207
;       REFERENCE/DOCKET NUMBER:  4834.US.P6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  708-937-6365
;       TELEFAX:  708-937-9556
;   INFORMATION FOR SEQ ID NO:  22:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  425 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-867-611-22

  Query Match             10.0%;  Score 77;  DB 3;  Length 425;
  Best Local Similarity   23.1%;  Pred. No. 3.5;
  Matches   30;  Conservative   19;  Mismatches   43;  Indels   38;  Gaps    4;

Qy         12 LEEEQYELGFKEGQIQGTKDQYLEGKE-----------YGYQTGFQRFLI--------- 50
              |: | | | |    |   :|:: || |           |||: || | |  :
Db        146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy         51 IGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKN--IKKAR 108
              |  :::|   |                   :   :||     |    || : :  |:::|
Db        206 IEMLEQLRVLWYG--------------EKIHVAVAQEVPGTGVDTPEDLDPSTRIRRSR 250

Qy        109 NKLRVIASIT 118
              |  :|| ::|
Db        251 NLGKVIDTLT 260


RESULT 15
PCT-US92-06965A-27
; Sequence 27, Application PC/TUS9206965A
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, S.
;    APPLICANT:  DESAI, S.
;    APPLICANT:  DAILEY, S.
;    TITLE OF INVENTION:  HCV SYNTHETIC PEPTIDE FROM NS1 REGION
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  U.S.
;      ZIP:  60065-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06965A
;      FILING DATE:  19920821
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834PC.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  425 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-06965A-27

  Query Match             10.0%;  Score 77;  DB 5;  Length 425;
  Best Local Similarity   23.1%;  Pred. No. 3.5;
  Matches   30;  Conservative   19;  Mismatches   43;  Indels   38;  Gaps    4;

Qy         12 LEEEQYELGFKEGQIQGTKDQYLEGKE-----------YGYQTGFQRFLI--------- 50
              |: | | | |    |   :|:: || |           |||: || | |  :
Db        146 LDAEGYALYFSRATIPWDRDRFAEGLETVGDNFLRHLGIYGYRAGFIRRYVNWQPSPLEH 205

Qy         51 IGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKN--IKKAR 108
              |  :::|   |                   :   :||     |    || : :  |:::|
Db        206 IEMLEQLRVLWYG--------------EKIHVAVAQEVPGTGVDTPEDLDPSTRIRRSR 250

Qy        109 NKLRVIASIT 118
              |  :|| ::|
Db        251 NLGKVIDTLT 260


Search completed: February 16, 2004, 14:15:31
Job time : 31 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 12:53:01 ; Search time 39 Seconds
(without alignments)
602.347 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	772	100.0	148	22	AAB30859	Amino acid sequenc
2	764	99.0	148	22	AAU15103	Protein encoded by
3	764	99.0	148	23	ABP73225	Candida albicans e
4	116	15.0	404	21	AAG30345	Arabidopsis thalia
5	99.5	12.9	237	22	ABG28105	Novel human diagno
6	99.5	12.9	293	22	ABB11923	Human GK003 homolo
7	99.5	12.9	293	22	AAM79451	Human protein SEQ
8	99.5	12.9	310	22	AAM78467	Human protein SEQ
9	99	12.8	107	21	AAG01966	Human secreted pro
10	95.5	12.4	141	22	ABB67857	Drosophila melanog
11	95	12.3	1646	23	ABP69013	Human polypeptide
12	90.5	11.7	525	23	AAU12041	Clostridium diffic
13	90	11.7	475	22	ABB68543	Drosophila melanog
14	87.5	11.3	803	23	AAU12040	Clostridium diffic
15	87	11.3	239	20	AAY20137	B. burgdorferi ant
16	87	11.3	274	20	AAY20136	B. burgdorferi ant
17	86	11.1	2285	20	AAW98149	Bacillus subtilis
18	85	11.0	507	21	AAB18173	Plasmodium falcipa
19	83.5	10.8	880	22	AAB96332	Putative P. abyssi
20	83.5	10.8	930	19	AAW55961	Human transient re
21	83.5	10.8	930	24	ABB99469	Amino acid sequenc
22	83.5	10.8	1211	22	ABB64663	Drosophila melanog
23	83	10.8	1184	22	AAU33733	Staphylococcus aur
24	83	10.8	1188	22	AAU36532	Staphylococcus aur
25	82.5	10.7	185	21	AAG54838	Arabidopsis thalia
26	82.5	10.7	435	23	ABP73423	Candida albicans e
27	82	10.6	724	22	AAG82994	S. epidermidis ope
28	82	10.6	756	23	ABP38708	Staphylococcus epi
29	82	10.6	1516	21	AAB18195	Plasmodium falcipa
30	81	10.5	320	22	AAU38192	Salmonella typhi c
31	81	10.5	440	22	AAU47019	Propionibacterium
32	81	10.5	470	23	ABP30175	Streptococcus poly
33	81	10.5	476	23	ABP26844	Streptococcus poly
34	81	10.5	521	20	AAY04366	Borrelia burgdorfe
35	81	10.5	699	22	ABB64367	Drosophila melanog
36	81	10.5	1164	19	AAW40541	Mutant C-beta prot
37	81	10.5	1188	21	AAB18183	Plasmodium falcipa
38	80.5	10.4	254	23	ABP73674	Candida albicans e
39	80	10.4	280	22	ABB58792	Drosophila melanog
40	80	10.4	1156	20	AAY34753	Amino acid sequenc
41	79.5	10.3	368	23	ABU52170	Helicobacter pylor
42	79.5	10.3	600	19	AAW98668	H. pylori GHPO 504
43	79.5	10.3	600	22	AAB68337	Amino acid sequenc
44	79.5	10.3	600	22	AAB68338	Amino acid sequenc
45	79.5	10.3	677	22	AAB68339	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB30859
ID AAB30859 standard; Protein; 148 AA.
XX
AC AAB30859;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a Candida albicans protein CaNL260.
XX
KW CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CaNL260; CaDR361;
KW antifungal; fungal infection; pathogenic fungi.
XX
OS Candida albicans.
XX
PN WO200075305-A2.
XX
PD 14-DEC-2000.
XX
PF 08-JUN-2000; 2000WO-FR01567.
XX
PR 09-JUN-1999; 99FR-0007250.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Lalanne J, Rocher C;
XX
DR WPI; 2001-050024/06.
DR N-PSDB; AAC86750.
XX
PT New polynucleotides from Candida albicans and their derived proteins,
PT useful for diagnosis and treatment of fungal infections and for drug

PT screening -
XX
PS Claim 11; Page 78-79; 89pp; French.
XX
CC The present sequence represents a Candida albicans protein. The
CC specification describes genes CaDRA472, CaDR489, 1CaDR527, 2CaDR527,
CC CaFLO24, CaNL260, and CaDR361. These genes are essential for
CC survival, and so are good targets for antifungal agents. The Candida
CC albicans genes and their derived proteins are used to screen compounds
CC for the ability to inhibit the activity of the protein, i.e. for
CC antifungal activity. The proteins are also used to generate a protective
CC antibody response against fungal infections in mammals. The Candida
CC albicans proteins and genes, and their antibodies, are used for
CC diagnosing fungal infections, specifically C. albicans (in standard
CC amplification, hybridisation or immunological assays, and for studying
CC pathogenic fungi.
XX
SQ Sequence 148 AA;

Query Match 100.0%; Score 772; DB 22; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.3e-71;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60
Db 1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60

Qy 61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120
Db 61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120

Qy 121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
Db 121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148

RESULT 2
AAU15103
ID AAU15103 standard; Protein; 148 AA.
XX
AC AAU15103;
XX
DT 04-DEC-2001 (first entry)
XX
DE Protein encoded by C. albicans essential gene CaYNL260C.
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection.
XX
OS Candida albicans.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05551.
XX
PR 18-FEB-2000; 2000US-0183534.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
DR N-PSDB; AAS23431.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes -
XX
PS Claim 43; Page 237-238; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans
CC essential genes.
XX
SQ Sequence 148 AA;

Query Match 99.0%; Score 764; DB 22; Length 148;
Best Local Similarity 98.6%; Pred. No. 1.5e-70;
Matches 146; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60
Db 1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60

Qy 61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120
Db 61 WLSHIDQYNNSSSLRNHLNNLENILAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120

Qy 121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
Db 121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148

RESULT 3
ABP73225
ID ABP73225 standard; Protein; 148 AA.
XX
AC ABP73225;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7062.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49486.
XX
PR 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
DR N-PSDB; ABZ31775.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression -
XX
PS Claim 44; SEQ ID NO 7062; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an

CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 148 AA;

Query Match 99.0%; Score 764; DB 23; Length 148;
Best Local Similarity 98.6%; Pred. No. 1.5e-70;
Matches 146; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSDIDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60

Qy         61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120
              ||||||||||||||||||||:|:||||||||||||||||||||||||||||||||||||||
Db         61 WLSHIDQYNNSSSLRNHLNNLENILAQISITNGDKEVEDYEKNIKKARNKLRVIASITKE 120

Qy        121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
              ||||||||||||||||||||||||||||
Db        121 TWKIDSLDNLVKEVGGTLQVSENPDDMW 148
```

RESULT 4
AAG30345
ID AAG30345 standard; Protein; 404 AA.
XX
AC AAG30345;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 36261.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 116; DB 21; Length 404;
Best Local Similarity 26.0%; Pred. No. 0.0022;
Matches 32; Conservative 22; Mismatches 51; Indels 18; Gaps 2;

```
Qy          6 IDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS-- 63
              :| :: |||   : || ||  :|      : :  | : ||:   :||: :     | |
Db         10 LDCIVRLEETHVQQGFDEGYEEGLVSGREDARHLGLKLGFETGELIGFYRGCSALWNSAL 69

Qy         64 HIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKN----------------IKKA 107
               ||     |  |  |||:   :: :| : : : | :|  |:                | :|
Db         70 RIDPTRFSPQLHKHLNDFHVLLDKIPLLDPEDEAKDGIKDDLREMLREMLGCCKVYISEA 129

Qy        108 RNK 110
               |||
Db        130 RNK 132
```

RESULT 5
ABG28105
ID ABG28105 standard; Protein; 237 AA.
XX
AC ABG28105;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #28096.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS92292.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 58464; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 237 AA;

Query Match 12.9%; Score 99.5; DB 22; Length 237;
Best Local Similarity 26.7%; Pred. No. 0.053;
Matches 32; Conservative 23; Mismatches 46; Indels 19; Gaps 3;

```
Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
Db        121 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 179

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
Db        180 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 221
```

RESULT 6
ABB11923
ID ABB11923 standard; peptide; 293 AA.
XX
AC ABB11923;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human GK003 homologue, SEQ ID NO:2293.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA09167.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 20; Page 277; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 293 AA;

Query Match 12.9%; Score 99.5; DB 22; Length 293;
Best Local Similarity 26.7%; Pred. No. 0.07;
Matches 32; Conservative 23; Mismatches 46; Indels 19; Gaps 3;

```
Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
Db        121 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 179

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
Db        180 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 221
```

```
RESULT 7
AAM79451
ID   AAM79451 standard; Protein; 293 AA.
XX
AC   AAM79451;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3097.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52584.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 251-252; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   293 AA;

  Query Match             12.9%;  Score 99.5;  DB 22;  Length 293;
  Best Local Similarity   26.7%;  Pred. No. 0.07;
  Matches   32;  Conservative   23;  Mismatches   46;  Indels   19;  Gaps    3;

Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
              |::    | |:::|   |       :|   ||: | :  |  | ::  :   ||    :||
Db        121 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 179

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              :|:| | :||| | : |           |:|           |: : |||  :  :| ||::
Db        180 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 221


RESULT 8
AAM78467
ID   AAM78467 standard; Protein; 310 AA.
XX
AC   AAM78467;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1129.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK51600.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 3362-3363; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   310 AA;

  Query Match             12.9%;  Score 99.5;  DB 22;  Length 310;
  Best Local Similarity   26.7%;  Pred. No. 0.076;
  Matches   32;  Conservative   23;  Mismatches   46;  Indels   19;  Gaps    3;

Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRPLIIGYIQELMKFWLSHIDQYNN 70
              |::    | |:::|   |       :|   ||: | :  |  | ::  :   ||    :||
Db        121 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 179
```

```
Qy          71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
               :|:| | :||| | : |           |:|          |: : |||  :  :| ||::
Db         180 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 221


RESULT 9
AAG01966
ID   AAG01966 standard; Protein; 107 AA.
XX
AC   AAG01966;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 6047.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01972.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 6047; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   107 AA;

  Query Match             12.8%;  Score 99;  DB 21;  Length 107;
  Best Local Similarity   27.0%;  Pred. No. 0.021;
  Matches   30;  Conservative   21;  Mismatches   42;  Indels   18;  Gaps    2;

Qy          20 GFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLN 79
               |:::|    |       :|    ||: | :   |  | ::  :    ||    :||:|:| | :|
Db           8 GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNNNSTLINKIN 67

Qy          80 NLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
               || | : |           |:|          |: : |||  :  :| ||::
Db          68 NLLDAVGQ---------CEEY---------VLKHLKSITPXSHVVDLLDSI 100


RESULT 10
ABB67857
ID   ABB67857 standard; Protein; 141 AA.
XX
AC   ABB67857;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 30363.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11960.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 30363; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   141 AA;

  Query Match             12.4%;  Score 95.5;  DB 22;  Length 141;
  Best Local Similarity   25.5%;  Pred. No. 0.068;
  Matches   28;  Conservative   29;  Mismatches   42;  Indels   11;  Gaps    5;

Qy           7 DNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHI- 65
               |:::  ||::  ||::|    | ||    :| | ||: |: : : :|  :|| |  |  :
Db          13 DDIVLTEEKEARLGYEE----GLKDGQEQGNEEGYKLGYAQGVSLG--EELGKI-LGQVV 65

Qy          66 --DQYNNSSSLRNHLNNLEDIMAQISITNGDK-EVEDYEKNIKKARNKLR 112
                 |  ::  :|  |  |  :: :   ||  : ::    ::|: :  :||
Db          66 AQQQLKHTDKVRRSLEQLRSLIEEFPRTNDPQADIVGAVQDIRSSHRRLR 115


RESULT 11
ABP69013
ID   ABP69013 standard; Protein; 1646 AA.
XX
AC   ABP69013;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1060.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
```

```
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   N-PSDB; ABZ11230.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
PS   Claim 9; SEQ ID NO 1060; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1646 AA;

  Query Match             12.3%;  Score 95;  DB 23;  Length 1646;
  Best Local Similarity   25.2%;  Pred. No. 2.1;
  Matches   36;  Conservative   27;  Mismatches   40;  Indels   40;  Gaps    7;

Qy         24 GQIQGTKDQYLEGKEYGY-----QTGFQRFLIIGYIQELMKFWLSH---------------- 64
              | :||  :: ||  :  :     || :|   :  | ||::: | :|
Db        626 GALQGKVEEDLELLDKSFETLADQTEWQSSHLFKYFQEVVQLWEAHQSELLVQELELEKR 685

Qy         65 IDQYNNSSSLRN-----HLNNLEDIMAQISITNGDKE--------VEDYEKNIKKARNKL 111
              ::|:    || :     ||: | | : | |    |||        |:|| ||:|
Db        686 MEQHRQKHSLESQVQEAHLDRLLDQLRQQS----DKETLAFHLEKVKDYLKNMKSRYECF 741

Qy        112 RVIASITKETWKIDSLDNLVKEV 134
                :  :|||  :  ::  ::||:
Db        742 HTL--LTKEVMEYPAI--MLKEL 760


RESULT 12
AAU12041
ID   AAU12041 standard; Protein; 525 AA.
XX
AC   AAU12041;
XX
DT   09-APR-2002  (first entry)
XX
DE   Clostridium difficile S-layer protein cell wall binding portion (ORF7).
XX
KW   Gene expression cassette; secretory leader sequence; anti-tumour therapy;
KW   clostridial N-acetylmuramoyl-L-alanine amidase-like protein; cytostatic;
KW   fibrinolysis; gangrenous tissue; necrotic tissue; infection;
KW   S-layer protein cell wall binding portion; ORF7.
XX
OS   Clostridium difficile.
XX
PN   WO200194599-A1.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-SE01280.
XX
PR   07-JUN-2000; 2000SE-0002139.
PR   26-APR-2001; 2001SE-0101479.
XX
PA   (SMIT-) SMITTSKYDDSINSTITUTET.
XX
PI   Burman LG,  Akerlund T,  Mukherjee K,  Katagihallimath N;
XX
DR   WPI; 2002-122121/16.
XX
PT   New gene expression cassette comprising a secretory leader sequence
PT   encoding a signal peptide from Clostridium difficile, useful e.g. for
PT   presenting polypeptides on bacterial cell surface, or as anti-tumour
PT   therapy  -
XX
PS   Claim 6; Page 77-79; 113pp; English.
XX
CC   The present invention relates to a new gene expression cassette
CC   comprising a secretory leader sequence encoding a signal peptide from
CC   Clostridium difficile or signal peptides of analogous exported
CC   clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked
CC   to a DNA sequence encoding a heterologous polypeptide. The gene
CC   expression cassette is useful for presenting polypeptides on the surface
CC   of bacterial cells, and/or secreting them into the surroundings of the
CC   cells, for mucosal immunisation, induction of immunological tolerance
CC   and anti-tumour therapy in humans and animals, as vaccines, and for the
CC   treatment of prevention of fibrinolysis in arterial or venous occlusion
CC   and/or revitalising gangrenous or necrotic tissue in various diseases.
CC   The gene expression cassette may also be used to produce in the gut,
CC   e.g. peptides and enzymes for therapy and prophylaxis of various
CC   diseases; single, fusion or multiple polypeptide antigens of microbial,
CC   animal or mammalian origin for neonatal immune balancing, vaccination
CC   against infections; and carrier molecules separated or fused to the
CC   antigen to amplify or modulate immune response. Spores produced by
CC   clostridia, e.g. Clostridium difficile are useful for both industrial
CC   production of vaccines and for local production of the desired
CC   polypeptides at the body sites desired. AAU12037-AAU12048 represent
CC   C. difficile S-layer protein cell wall binding portions encoded by
CC   ORF1, ORF3, ORF5-7, ORF9, ORF11, ORFD, ORFE, ORFG, ORFH or ORFI
CC   respectively. The DNAs encoding these portions may be used in a
CC   gene expression cassette of the present invention.
XX
SQ   Sequence   525 AA;

  Query Match             11.7%;  Score 90.5;  DB 23;  Length 525;
  Best Local Similarity   24.3%;  Pred. No. 1.3;
  Matches   34;  Conservative   26;  Mismatches   43;  Indels   37;  Gaps    6;

Qy         22 KEGQIQGTKDQYL----------EGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
              || :  | || ||          | |  |     :|             ::  ::|
Db         95 KEIKRLGAKDIYLIGGTAVLNKDIENKLKGNGLNVER-------------INGKNRYET 140

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVED-YEKNIKKARNKLRVIASITKETWKIDSLDN 129
```

```
                |  | | | :::||  :::: ||:| : |        |:||: :| |  |:   ::: |
Db        141 SLILANKLKDIKDI-KEVAVVNGEKGLSDAVSVGAPAAQNKMPIILSNPKD--GVEAFDK 197

Qy        130 LVKE--------VGGTLQVS 141
                :::        :|||  ||
Db        198 FIRDEKVIKAYVIGGTNSVS 217


RESULT 13
ABB68543
ID   ABB68543 standard; Protein; 475 AA.
XX
AC   ABB68543;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 32421.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL12646.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 32421; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   475 AA;

  Query Match           11.7%;  Score 90;  DB 22;  Length 475;
  Best Local Similarity   25.7%;  Pred. No. 1.3;
  Matches   37;  Conservative   32;  Mismatches   51;  Indels   24;  Gaps    7;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTK--DQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              |  :   |::  :   | | | |:  :|  |      :  |       ||||
Db         47 DTAIRQAEKEKTILIDELQHQNTRLTEQIQEAHATELKLSAQ-------IQELK------ 93

Qy         65 IDQYN-NSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKAR-NKLRVIASITKETW 122
                |||:  :|||: |:|:|| |  ::::| | :  :: |: :: |:  |  : :|: : :
Db         94 -DQYHYRNSSLQEHVNSLESIKTELNLTTGKR--QELERRLQIAQEEKESLTSSLEESSD 150

Qy        123 KIDSLDNLVKE----VGGTLQVSE 142
               :|  |:   :|    :  |||  |

Db        151 RIHMLERHAREQETKLETTLQALE 174


RESULT 14
AAU12040
ID   AAU12040 standard; Protein; 803 AA.
XX
AC   AAU12040;
XX
DT   09-APR-2002  (first entry)
XX
DE   Clostridium difficile S-layer protein cell wall binding portion (ORF6).
XX
KW   Gene expression cassette; secretory leader sequence; anti-tumour therapy;
KW   clostridial N-acetylmuramoyl-L-alanine amidase-like protein; cytostatic;
KW   fibrinolysis; gangrenous tissue; necrotic tissue; infection;
KW   S-layer protein cell wall binding portion; ORF6.
XX
OS   Clostridium difficile.
XX
PN   WO200194599-A1.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-SE01280.
XX
PR   07-JUN-2000; 2000SE-0002139.
PR   26-APR-2001; 2001SE-0101479.
XX
PA   (SMIT-) SMITTSKYDDSINSTITUTET.
XX
PI   Burman LG,  Akerlund T,  Mukherjee K,  Katagihallimath N;
XX
DR   WPI; 2002-122121/16.
XX
PT   New gene expression cassette comprising a secretory leader sequence
PT   encoding a signal peptide from Clostridium difficile, useful e.g. for
PT   presenting polypeptides on bacterial cell surface, or as anti-tumour
PT   therapy  -
XX
PS   Claim 6; Page 75-77; 113pp; English.
XX
CC   The present invention relates to a new gene expression cassette
CC   comprising a secretory leader sequence encoding a signal peptide from
CC   Clostridium difficile or signal peptides of analogous exported
CC   clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked
CC   to a DNA sequence encoding a heterologous polypeptide. The gene
CC   expression cassette is useful for presenting polypeptides on the surface
CC   of bacterial cells, and/or secreting them into the surroundings of the
CC   cells, for mucosal immunisation, induction of immunological tolerance
CC   and anti-tumour therapy in humans and animals, as vaccines, and for the
CC   treatment of prevention of fibrinolysis in arterial or venous occlusion
CC   and/or revitalising gangrenous or necrotic tissue in various diseases.
CC   The gene expression cassette may also be used to produce in the gut,
CC   e.g. peptides and enzymes for therapy and prophylaxis of various
CC   diseases; single, fusion or multiple polypeptide antigens of microbial,
CC   animal or mammalian origin for neonatal immune balancing, vaccination
CC   against infections; and carrier molecules separated or fused to the
CC   antigen to amplify or modulate immune response. Spores produced by
CC   clostridia, e.g. Clostridium difficile are useful for both industrial
CC   production of vaccines and for local production of the desired
CC   polypeptides at the body sites desired. AAU12037-AAU12048 represent
CC   C. difficile S-layer protein cell wall binding portions encoded by
CC   ORF1, ORF3, ORF5-7, ORF9, ORF11, ORFD, ORFE, ORFG, ORFH or ORFI
CC   respectively. The DNAs encoding these portions may be used in a
CC   gene expression cassette of the present invention.
XX
SQ   Sequence   803 AA;

  Query Match           11.3%;  Score 87.5;  DB 23;  Length 803;
  Best Local Similarity   27.5%;  Pred. No. 4.7;
  Matches   36;  Conservative   20;  Mismatches   58;  Indels   17;  Gaps    6;
```

```
Qy          23 EGQIQGTKDQYLEGKEYGYQTGFQRFLI-IGYIQELMKFWLSHIDQYNNSSSLRNHLNNL 81
Db         570 ELQRLGTKKVYLIGGENSLSKNVQTQLSNMGISVER----ISGSDRYKTSISLAQKLNSI 625

Qy          82 EDIMAQISITNGDKEVED-YEKNIKKARNKLRVIASITKETWKIDSLDNLVKE------- 133
Db         626 KSV-SQVAVANGVNGLADAISVGAAAADNNMPII--LTNEKSELQGADEFLNSSKITKSY 682

Qy         134 -VGGTLQVSEN 143
Db         683 IIGGTATLSSN 693
```

RESULT 15
AAY20137
ID AAY20137 standard; Protein; 239 AA.
XX
AC AAY20137;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, t02A.aa BB002.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61834.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 12; Page 211; 275pp; English.
XX
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 239 AA;

  Query Match   11.3%; Score 87; DB 20; Length 239;
  Best Local Similarity   23.5%; Pred. No. 1;
  Matches   42; Conservative   32; Mismatches   47; Indels   58; Gaps   8;

```
Qy           7 DNVLNLEEEQYELGFKEGQIQGTKDQYLE--GKE-------YGYQ----TGFQRFLIIGY 53
Db          31 ENIQNFKDKSGDLGASDEKFMGTTASELKAIGKELEDRKNQYDIQIAKITNEESNLLDTY 90

Qy          54 IQE-----------LMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEK 102
Db          91 IRAYELANENEKMLLKRFLLSSLDY------KKENIETLKEIL--------EKLINNYEN 136

Qy         103 NIKKARNKLRVIA------------------SITKETWKIDSLDNLVKEVGGTLQVSE 142
Db         137 DPKIAANFLYRIALDIQLKLEKHLKSINEKLDTLSKENSKED-LEALLEQVKSALQLQE 194
```

Search completed: February 16, 2004, 14:09:50
Job time : 40 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 12:08:41 ; Search time 328 Seconds
(without alignments)
5020.073 Million cell updates/sec

Title: US-09-980-054A-11
Perfect score: 447
Sequence: 1 atgtcagatatagatataga.....accccgatgatatgtggtga 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	442.2	98.9	447	13	US-10-032-585-6062	Sequence 6062, Ap
c	2	70.4	15.7	4985	13	US-10-056-405-10	Sequence 10, Appl
c	3	70.4	15.7	4985	15	US-10-094-240-10	Sequence 10, Appl
c	4	65	14.5	65	13	US-10-032-585-2884	Sequence 2884, Ap
	5	64.2	14.4	3673778	13	US-10-312-841-2	Sequence 2, Appli
c	6	62.6	14.0	9539	13	US-10-240-453-53	Sequence 53, Appl
c	7	62.6	14.0	9539	15	US-10-239-676-51	Sequence 51, Appl
	8	61.2	13.7	168575	13	US-10-178-194-1	Sequence 1, Appli
c	9	60.2	13.5	17183	13	US-10-311-455-460	Sequence 460, App
c	10	59.8	13.4	6109	13	US-10-311-455-299	Sequence 299, App
c	11	59.8	13.4	12237	13	US-10-311-455-2331	Sequence 2331, Ap
c	12	59.4	13.3	446	10	US-09-960-352-3400	Sequence 3400, Ap
c	13	59.2	13.2	12142	13	US-10-311-455-1646	Sequence 1646, Ap
c	14	58.2	13.0	3673778	13	US-10-312-841-2	Sequence 2, Appli
c	15	57.8	12.9	516	10	US-09-960-352-5785	Sequence 5785, Ap
c	16	57.6	12.9	9539	13	US-10-240-453-54	Sequence 54, Appl
c	17	57.6	12.9	9539	15	US-10-239-676-52	Sequence 52, Appl
c	18	57.4	12.8	6641	13	US-10-311-455-287	Sequence 287, App
c	19	57.4	12.8	6641	13	US-10-240-452-35	Sequence 35, Appl
c	20	57	12.8	653	13	US-10-027-632-113792	Sequence 113792,
c	21	57	12.8	653	14	US-10-027-632-113792	Sequence 113792,
c	22	56.8	12.7	17131	13	US-10-311-455-1026	Sequence 1026, Ap
c	23	56.4	12.6	6301	13	US-10-311-455-26	Sequence 26, Appl
c	24	56.4	12.6	3673778	13	US-10-312-841-1	Sequence 1, Appli
c	25	55.6	12.4	9963	13	US-10-311-455-667	Sequence 667, App
c	26	55.2	12.3	6175	13	US-10-311-455-1280	Sequence 1280, Ap
c	27	55	12.3	55	13	US-10-032-585-3062	Sequence 3062, Ap
c	28	54.8	12.3	419	10	US-09-960-352-11234	Sequence 11234, A
c	29	54.8	12.3	15548	13	US-10-311-455-2128	Sequence 2128, Ap
c	30	54.6	12.2	14006	13	US-10-311-455-1931	Sequence 1931, Ap
c	31	54.6	12.2	53332	10	US-09-801-861-3	Sequence 3, Appli
c	32	54.6	12.2	53332	15	US-10-224-562-3	Sequence 3, Appli
	33	54.4	12.2	5770	13	US-10-311-455-14	Sequence 14, Appl
c	34	54.2	12.1	6113	13	US-10-311-455-1796	Sequence 1796, Ap
c	35	54.2	12.1	6535	13	US-10-311-455-909	Sequence 909, App
c	36	54.2	12.1	7346	13	US-10-311-455-318	Sequence 318, App
c	37	54.2	12.1	7597	13	US-10-311-455-986	Sequence 986, App
	38	54.2	12.1	15732	13	US-10-240-453-108	Sequence 108, App
	39	54.2	12.1	15732	15	US-10-239-676-96	Sequence 96, Appl
	40	54	12.1	335913	11	US-09-754-853A-2	Sequence 2, Appli
	41	54	12.1	335913	11	US-09-754-853A-3	Sequence 3, Appli
c	42	53.8	12.0	6167	13	US-10-240-453-243	Sequence 243, App
c	43	53.8	12.0	13584	13	US-10-311-455-588	Sequence 588, App
c	44	53.6	12.0	393	10	US-09-960-352-4582	Sequence 4582, Ap
	45	53.6	12.0	2437	13	US-10-311-455-2324	Sequence 2324, Ap

ALIGNMENTS

RESULT 1
US-10-032-585-6062
; Sequence 6062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6062
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6062

Query Match 98.9%; Score 442.2; DB 13; Length 447;
Best Local Similarity 99.3%; Pred. No. 4.1e-67;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
```

```
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447

Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447
```

RESULT 2
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match 15.7%; Score 70.4; DB 13; Length 4985;
Best Local Similarity 51.0%; Pred. No. 0.0073;
Matches 196; Conservative 0; Mismatches 181; Indels 7; Gaps 1;

```
Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63

Db       2850 TAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAATAAGAATAATAATAA 2791

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123

Db       2790 TAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTATTATTATGAT 2731

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183

Db       2730 AATAATGTATAATAATGAATAACAATA-------ATAATAATTAATAATAATAAATTAAT 2678

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243

Db       2677 TAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAATAATAATAATAATAAT 2618

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303

Db       2617 AATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATAATAATAAT 2558

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363

Db       2557 AATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATAATAAT 2498

Qy        364 TGGAAAATTGATTCATTGGATAAT 387

Db       2497 AATAATAATAATAATATTAATAAT 2474
```

RESULT 3
US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 15.7%; Score 70.4; DB 15; Length 4985;
Best Local Similarity 51.0%; Pred. No. 0.0073;
Matches 196; Conservative 0; Mismatches 181; Indels 7; Gaps 1;

```
Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63

Db       2850 TAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAATAAGAATAATAATAA 2791

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123

Db       2790 TAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTATTATTATGAT 2731

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183

Db       2730 AATAATGTATAATAATGAATAACAATA-------ATAATAATTAATAATAATAAATTAAT 2678

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243

Db       2677 TAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAATAATAATAATAATAAT 2618

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303

Db       2617 AATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATAATAATAAT 2558

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363

Db       2557 AATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATAATAAT 2498

Qy        364 TGGAAAATTGATTCATTGGATAAT 387

Db       2497 AATAATAATAATAATATTAATAAT 2474
```

RESULT 4
US-10-032-585-2884/c
; Sequence 2884, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2884
; LENGTH: 65

```
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-032-585-2884

  Query Match             14.5%;  Score 65;  DB 13;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   65;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 6

Qy         64 AAAGA 68
              |||||
Db          5 AAAGA 1


RESULT 5
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             14.4%;  Score 64.2;  DB 13;  Length 3673778;
  Best Local Similarity   54.2%;  Pred. No. 0.57;
  Matches  173;  Conservative    0;  Mismatches  143;  Indels    3;  Gaps    2;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              || | | | |||| ||    | |  | || ||||  ||| | || ||   | | |   |
Db     554727 ATATTAAAATTAGAAATTTTAATTAAAATATATTTTAAAGGAAAAAAAGTTATAATTTAA 554786

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAG-GAAAAGAATATGG 119
              || |    | | ||| ||   | | ||| |||||||||||     |||||||
Db     554787 ATTTATTTATATTATATATAATTTATATAGGATTAATATTTAGAATATTTAAGAATAGTT 554846

Qy        120 TTATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATT 179
              ||||   |    || || ||| |  ||  | | |||||  | | ||  || |||
Db     554847 TTATAGGAAAATGATTAAAAGATGTGAAAGGGTAGTTTATAAAAAGGGATATATAAAAGA 554906

Qy        180 TTGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAA 239
              ||  | | |  ||| |  | ||  | ||||   || ||           ||| |||||
Db     554907 TTAATAAACGTATAAAAGTTAAGTTTATAATAGTTATTAGAGAATTGTGAATTAGAATAG 554966

Qy        240 TTTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTA 299
              |||   | ||| |||     ||| || |  ||   ||  | ||  | ||||| || ||||
Db     554967 TTTTTGAAATAATAT--TTTAAAATTATTAAATAGATAAATATTGAAAAGTTTAATATTA 555024

Qy        300 TGAAAAAAATATTAAAAAG 318
              | ||||||   || |  | |
Db     555025 TTAAAAATTAATGAGGATG 555043


RESULT 6
US-10-240-453-53/c
; Sequence 53, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associate
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 53
;   LENGTH: 9539
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-53

  Query Match             14.0%;  Score 62.6;  DB 13;  Length 9539;
  Best Local Similarity   47.8%;  Pred. No. 0.19;
  Matches  182;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              ||| | | |  || |     |   || ||||||| || || |  ||||| ||  |   | |
Db       6192 TCAAAAAAATTTAAAATATAACCATAAAATTTATAAAAAAACAAATATAAACGAAATTAT 6133

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
              ||| |     |   |  ||  || | ||  |||| |  ||||  | || | |
Db       6132 AAAACAAAAACGAACTTAAAAATTCAAAAACACATTTTAAACTCAAAAATAAATTATAAA 6073

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
              | |   ||  | ||  | |  ||  |    | | | |||| ||| ||   ||
Db       6072 AACAAACGAAATTAAAAAATCGAAACGAAATTTAAAAATCAAAAATAAAATTAAAAAATA 6013

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              |||   |  | ||     | ||| |              | |||       |  |
Db       6012 ATATAAAAATCAAATTCTAAAAATCATAAAAAAATCTAAAAAATTCAAAAATTTTTCTAA 5953

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              || | | ||       |||  || | ||   ||  | | ||| || |  | | | || ||
Db       5952 AAAAAAAATAAAATTTAAAAATCAAAAATCTATTTAAACAAACAAAATATAACTAATAAA 5893

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              |||||  |||||||  ||| | | || | ||     | |       || |||||| ||| |
Db       5892 AAAAAATTTAAAAAACCAAAATAAAATTAAAATAAAAAAAAACTTAAAACTAAAAAAATT 5833

Qy        364 TGGAAAATTGATTCATTGGAT 384
              ||||   ||  | |  ||
Db       5832 CAAAAAAACAATAAAATATAT 5812


RESULT 7
US-10-239-676-51/c
; Sequence 51, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
```

```
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
;   LENGTH: 9539
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51

  Query Match             14.0%;  Score 62.6;  DB 15;  Length 9539;
  Best Local Similarity   47.8%;  Pred. No. 0.19;
  Matches  182;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              ||| | | |  || |    |  || ||||||| || || |  ||||| ||  |   | |
Db       6192 TCAAAAAAATTTAAAATATAACCATAAAATTTATAAAAAAACAAATATAAACGAAATTAT 6133

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
              ||| |       |  ||  || | ||  |||| |   |||| | || | | |
Db       6132 AAAACAAAAACGAACTTAAAAATTCAAAAACACATTTTAAACTCAAAAATAAATTATAAA 6073

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
              | |   ||  | ||  | |   ||  |    | | | |||| ||| ||   ||
Db       6072 AACAAACGAAATTAAAAAATCGAAACGAAATTTAAAAATCAAAAATAAAATTAAAAAATA 6013

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              |||   |   | ||   | |||  |              | |||     |   |
Db       6012 ATATAAAAATCAAATTCTAAAAATCATAAAAAAATCTAAAAAATTCAAAAATTTTTCTAA 5953

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              || | | ||      |||  ||  | ||  ||  | | ||| ||  |  |  | || ||
Db       5952 AAAAAAAATAAAATTTAAAAATCAAAAATCTATTTAAACAAACAAAATATAACTAATAAA 5893

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              |||||  |||||||  ||| | | || | ||      | |      || |||||| ||| |
Db       5892 AAAAAATTTAAAAAACCAAAATAAAATTAAAATAAAAAAAAACTTAAAACTAAAAAAATT 5833

Qy        364 TGGAAAATTGATTCATTGGAT 384
                ||||   ||  | |  ||
Db       5832 CAAAAAAACAATAAAATATAT 5812


RESULT 8
US-10-178-194-1
; Sequence 1, Application US/10178194
; Publication No. US20030166904A1
; GENERAL INFORMATION:
;  APPLICANT: Berglind Ran Olafsdottir
;  APPLICANT:  Jeffrey Gulcher
;  TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
;  FILE REFERENCE: 2345.2001-000
;  CURRENT APPLICATION NUMBER: US/10/178,194
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: US/09/426,290
;  PRIOR FILING DATE: 1999-10-25
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 168575
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (21181)...(21403)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (95252)...(95430)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101753)...(101996)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (110324)...(110439)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (124058)...(124278)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (127009)...(127130)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128910)...(129139)
US-10-178-194-1

  Query Match             13.7%;  Score 61.2;  DB 13;  Length 168575;
  Best Local Similarity   48.1%;  Pred. No. 0.77;
  Matches  203;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATAT-GAATTAGGATTTAAA 66
              ||||| | ||| ||||||||||| | ||| |    |  |  |||| || ||      | |
Db     108964 ATATATAAATATATAATGTATTATAGTTATATATAATATTATATATAAATATATAATGTA 109023

Qy         67 GAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAA 126
              |  |   ||| |    |  | | || ||||| || |  | | |      |  | |  |
Db     109024 TTATATGTTATATATAATATTACATATAAATATCTATTAATATATATTATTTATATTATA 109083

Qy        127 ACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTA 186
              ||  || | |  || |   |  |    |||||  |  || | | || ||   ||
Db     109084 TCTAAATATATAATATATAACTTATTATTATATATTATAATATAACTTATATATTATATA 109143

Qy        187 TCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAA 246
              |  |||||  |  | |||  |  | || |  | |    | | || |  | |     |
Db     109144 TAATATATATAATATAATATAACTTATTATATATTATATATTTATATATAAATAAAATAT 109203

Qy        247 GATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAA 306
              |   |      | | ||  |||| | |||   ||| | ||  |  | | |||  ||| |
Db     109204 GTACTATATTAATATATGAATATATCTAATATTAATATACAATATATAAATTTATAAATA 109263

Qy        307 AATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGG 366
              ||| | |  |   | | |||| ||  |  |||  |  | ||  ||    | ||
Db     109264 TATAATGATTATATATTATATAATATATAAAATATATATTATGTAGGGAATCTGAATTTA 109323

Qy        367 AAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAA 426
              ||  ||| ||      || |  | | | | ||  |  |  | ||  || ||| |||
Db     109324 TTTATGTATTTATGTACATATATAAGGTAGGGAATATATATATATGTATTAGGTAGGGAA 109383

Qy        427 AA 428
              ||
Db     109384 TA 109385


RESULT 9
US-10-311-455-460/c
; Sequence 460, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
```

```
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 460
;   LENGTH: 17183
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-460

  Query Match             13.5%;  Score 60.2;  DB 13;  Length 17183;
  Best Local Similarity   49.9%;  Pred. No. 0.59;
  Matches  185;  Conservative    0;  Mismatches  178;  Indels    8;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
Db       4235 ATAAAAAAATAAATCATAAATACTACCTAAAAAAATTAAAAAACTATAAACCCTCTAAAC 4176

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
Db       4175 AAAAAAAACAACATTTACAAAAACATAAATATACAAAAACATAATATCTCTAAAAAAAAT 4116

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187
Db       4115 CAATATTATTAAAATATAAACTACATAATAAAAAATAATAATTAATTAAATAAATTAAAA 4056

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247
Db       4055 CTTACTTATAAAAAAATATATACATTTTTTCCCTATAAACAATAATAAATCATAATAAAT 3996

Qy        248 ATATTATGGCACAAATTTCTATAAC--------GAATGGAGATAAAGAAGTTGAAGATTA 299
Db       3995 TTTTAAATATACAAAATAACATAACCACATTTATATTTAAAAAAATAAATATACAATAA 3936

Qy        300 TGAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGA 359
Db       3935 TATAAAAAAATATACCAACCCAAAAAAAAACAATTAAAAAAATAAATTATATCATTCAATA 3876

Qy        360 AACTTGGAAAA 370
Db       3875 ATCCAATAAAA 3865


RESULT 10
US-10-311-455-299/c
; Sequence 299, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 299
;   LENGTH: 6109
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 214, 2796..2797, 4347
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-299

  Query Match             13.4%;  Score 59.8;  DB 13;  Length 6109;
  Best Local Similarity   46.3%;  Pred. No. 0.52;
  Matches  196;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       2081 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 2022

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       2021 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAACGAAAAAATAAATAAATAAAT 1962

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       1961 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAAATAAATAAATAAAT 1902

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       1901 AAATAAATAAATAAATAAATATATAAATAAATAAATAATAAATAAATATATAAATAAATA 1842

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db       1841 AATAAATAAATAAATAAATATATAAATAAATAAATAAAAAATAAAAAAATAAATAAATAA 1782

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db       1781 ATAAATTAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 1722

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGA 425
Db       1721 ATAAATAAATAAATAAATAAATAAATCGATATATAAATAAATAAATAAATAAATAAATAA 1662

Qy        426 AAA 428
Db       1661 ATA 1659


RESULT 11
US-10-311-455-2331/c
; Sequence 2331, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2331
;   LENGTH: 12237
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2331

  Query Match             13.4%;  Score 59.8;  DB 13;  Length 12237;
  Best Local Similarity   46.3%;  Pred. No. 0.63;
```

```
  Matches  196;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       6985 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAAAAAAAA 6926

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       6925 AAAAAAATAAATAAATCAACAAATAAATAAATAAATAAATAAATAAATAAAAAAAAAAAA 6866

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       6865 AAAAAAATAAATAAATCAATAAATAAATAAATAAAATAAATAAATAAATAAATAAATAAA 6806

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       6805 AAACAAAAAAAAAAATAAATAAATCAATAAATAAATAAATAAAATAAATAAATAAATAAA 6746

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db       6745 TAAATAAATAAATAAATAAAAAAAAAAAAAAAAAAATAAATAAATCAACAAATAAATAAA 6686

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db       6685 TAAATAAATAAATAAATAAAAAAAATAAATAAATAAATCAAAAATAAATAAATAAAAAAA 6626

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGA 425
Db       6625 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6566

Qy        426 AAA 428
Db       6565 ATA 6563
```

RESULT 12
US-09-960-352-3400/c
; Sequence 3400, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3400
;   LENGTH: 446
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

```
  Query Match             13.3%;  Score 59.4;  DB 10;  Length 446;
  Best Local Similarity   47.7%;  Pred. No. 0.29;
  Matches  174;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db        437 AAAAAAAAAAAAAAAAAAAAATAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAA 378

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db        377 AAAAAAAATAAAATAATATAAAAAATAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAA 318

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db        317 AATAAAAAAATAAAAAAAAAAATAAATAAAATAAATAAAAAAAATAAAAAAAAAAATAAAA 258

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db        257 AAAAAATAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 198

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db        197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 138

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db        137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 78

Qy        366 GAAAA 370
Db         77 AAAAA 73
```

RESULT 13
US-10-311-455-1646/c
; Sequence 1646, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1646
;   LENGTH: 12142
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1646

```
  Query Match             13.2%;  Score 59.2;  DB 13;  Length 12142;
  Best Local Similarity   48.1%;  Pred. No. 0.8;
  Matches  198;  Conservative    0;  Mismatches  213;  Indels    1;  Gaps    1;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       3204 AAATCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 3145

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       3144 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAAATAAAT 3085

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       3084 AAATAAAAAAAAAAATAAAAAAAAAATAATAAAAAATAAATAAATAAATAAAATAAAAAA 3025

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       3024 AAACGAAAAATTAACAAATAATAAATAAAAAAAAAAAAATAAATAAATAAATAAATAAAA 2965

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGA-GATAAAGAAGTTGAAGATTATGAAA 304
Db       2964 ATAAATAAAAATTAAACAAAAAATAACAAATAAACGAAAAATAAATAAAAAATTAATAAA 2905

Qy        305 AAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTT 364
Db       2904 AATAAATAAACGAATAAATAAATAAATAAATAAATACATTAAAAAAAAATAAATAAACTA 2845
```

```
Qy        365 GGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACA 416
Db       2844 AAACTAATTTCATTCAATAAATATTAATTAAATATCCACCTTATACTTAATA 2793
```

RESULT 14
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

```
  Query Match             13.0%;  Score 58.2;  DB 13;  Length 3673778;
  Best Local Similarity   48.8%;  Pred. No. 6.1;
  Matches  185;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
Db    1278754 ATTAAATACAAATATAAAACAAACAAAATAAAATTTAACCAAAAATAAAATATACCAAAA 1278695

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
Db    1278694 AAAAAAAAACCAAAAATACAAAATAAAAACTAAAAAACGATTACCGTATTAAACCTTAAA 1278635

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
Db    1278634 ATCTAAACTAAATAATAAAAAAAATAAAAAATTAAAATCAAACAATAAAAAATAAATTAA 1278575

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
Db    1278574 TAAATAAAAAACCCAAATAAAATTAAAAAATTTT-TAAAATAAAAATACTAAAAAACAAC 1278516

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
Db    1278515 ATAAAAAAACTAAAAATAAAATTTTAAAAAAACAATAAATATTAATAACAACCAAATTAA 1278456

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
Db    1278455 AAATATAACATAAAACGAAAAAAAAACAAAAAAAAATAAATCACTACAAAAAAAAAACAA 1278396

Qy        361 ACTTGGAAAATTGATTCAT 379
Db    1278395 AATACTAAAAAATCTACAT 1278377
```

RESULT 15
US-09-960-352-5785/c
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76),(90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785

```
  Query Match             12.9%;  Score 57.8;  DB 10;  Length 516;
  Best Local Similarity   47.6%;  Pred. No. 0.56;
  Matches  170;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy         30 AAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTCAAATACAAGGAACAAA 89
Db        514 ATAATAAAAAAATAAAAAAAAATATTAAATATAAAAAAAAAAATAAATAAATAATAAAAA 455

Qy         90 AGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTTCAACGATTTTTAAT 149
Db        454 AAAATAATAATAAATAAAAAATAAAAAATAAAAAAAAAAAAAATAAAAAATAAAAAAAAA 395

Qy        150 CATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATAGATCAATATAATAA 209
Db        394 AAAAAAAAAAATATAAAAAATAATAAAAAAAAAAAAAATAAATAAAAACAAAAATTATAA 335

Qy        210 CTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATGGCACAAATTTCTAT 269
Db        334 AAAAAAAAAAAAAATAAAAAAAAAAAATAAAATAAAAAAAAAAAAAAAATAAAAATATAA 275

Qy        270 AACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAAAAGGCAAGAAATAA 329
Db        274 AAATAAAAAAAAAAAAATAAAAAAAAATAATAAAAAATAAAAAAAAAAAAATAAAAAAAA 215

Qy        330 ATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTGATTCATTGGATAA 386
Db        214 ATTAAAAAAAAAAAAAATAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAATATAA 158
```

Search completed: February 16, 2004, 14:07:50
Job time : 352 secs

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 10:57:21 ; Search time 64 Seconds
(without alignments)
3082.785 Million cell updates/sec

Title: US-09-980-054A-11
Perfect score: 447
Sequence: 1 atgtcagatatagatataga..........accccgatgatatgtggtga 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	61.2	13.7	168575	4	US-09-426-290-1	Sequence 1, Appli
c	2	58.2	13.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	3	55.8	12.5	615	3	US-08-998-416-186	Sequence 186, App
	4	55.6	12.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	5	54.6	12.2	53332	4	US-09-801-861-3	Sequence 3, Appli
c	6	54.4	12.2	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	7	53.8	12.0	5340	4	US-09-627-122-21	Sequence 21, Appl
c	8	52.4	11.7	837	3	US-08-998-416-288	Sequence 288, App
c	9	51.2	11.5	636	3	US-08-998-416-1137	Sequence 1137, Ap
	10	49.8	11.1	6124	4	US-08-213-419B-3	Sequence 3, Appli
c	11	49	11.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	12	48.2	10.8	2427	4	US-09-601-198-70	Sequence 70, Appl
c	13	48	10.7	8920	2	US-08-446-855A-1	Sequence 1, Appli
c	14	48	10.7	8920	3	US-09-150-741-1	Sequence 1, Appli
c	15	47.2	10.6	5652	4	US-09-601-198-75	Sequence 75, Appl
	16	47	10.5	1665	4	US-09-601-198-104	Sequence 104, App
	17	46.4	10.4	7101	1	US-08-480-604A-9	Sequence 9, Appli
	18	46.4	10.4	7101	2	US-08-405-496A-9	Sequence 9, Appli
	19	46.4	10.4	7101	3	US-08-915-136-9	Sequence 9, Appli
	20	46.4	10.4	7101	4	US-08-957-310-9	Sequence 9, Appli
	21	46.4	10.4	7101	4	US-10-011-366-9	Sequence 9, Appli
	22	45.8	10.2	1275	4	US-09-134-001C-2113	Sequence 2113, Ap
	23	45.6	10.2	3095	6	5231168-1	Patent No. 5231168
	24	45.4	10.2	2435	3	US-09-306-593-1	Sequence 1, Appli
	25	45.2	10.1	4766	5	PCT-US93-07261-10	Sequence 10, Appl
c	26	45.2	10.1	580073	4	US-08-545-528D-1	Sequence 1, Appli
c	27	45	10.1	6768	1	US-08-107-755A-1	Sequence 1, Appli
c	28	45	10.1	8457	1	US-07-991-867B-1	Sequence 1, Appli
c	29	45	10.1	8457	2	US-08-544-332-1	Sequence 1, Appli
c	30	45	10.1	8457	4	US-09-370-861A-1	Sequence 1, Appli
	31	44.6	10.0	927	4	US-09-134-001C-150	Sequence 150, App
c	32	44.4	9.9	1689	1	US-07-991-867B-41	Sequence 41, Appl
c	33	44.4	9.9	1689	2	US-08-544-332-41	Sequence 41, Appl
c	34	44.4	9.9	1689	4	US-09-370-861A-41	Sequence 41, Appl
c	35	44.4	9.9	1947	4	US-09-370-861A-74	Sequence 74, Appl
c	36	44.4	9.9	4163	4	US-09-004-838-70	Sequence 70, Appl
c	37	44.4	9.9	4208	4	US-09-004-838-1	Sequence 1, Appli
	38	44.2	9.9	375	4	US-09-134-001C-2120	Sequence 2120, Ap
	39	44.2	9.9	2411	4	US-09-446-301A-15	Sequence 15, Appl
	40	44.2	9.9	2411	4	US-09-099-932-47	Sequence 47, Appl
	41	44.2	9.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	42	44	9.8	642	1	US-08-764-100-13	Sequence 13, Appl
c	43	44	9.8	643	1	US-08-764-100-7	Sequence 7, Appli
c	44	44	9.8	2993	1	US-08-764-100-2	Sequence 2, Appli
	45	44	9.8	2993	1	US-08-764-100-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

```
  Query Match             13.7%;  Score 61.2;  DB 4;  Length 168575;
  Best Local Similarity   48.1%;  Pred. No. 0.00042;
  Matches  203;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATAT-GAATTAGGATTTAAA 66
              |||||  | |||  ||||||||||| |||  |    |   ||||  || ||
Db     108964 ATATATAAATATATAATGTATTATAGTTATATATAATATTATATATAAATATATAATGTA 109023

Qy         67 GAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAA 126
                |  |   |||  |     |  |  | ||  |||||  ||   |   | |      |  |  |
Db     109024 TTATATGTTATATATAATATTACATATAAATATCTATTAATATATATTATTTATATTATA 109083

Qy        127 ACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTA 186
                ||  ||  |   |  ||  |     |   |   |||||  |   ||  |  |  |  | ||  ||     ||
Db     109084 TCTAAATATATAATATATAACTTATTATTATATATTATAATATAACTTATATATTATATA 109143
```

```
Qy        187 TCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAA 246
              |   ||||| |  |  ||| |  | || |  | |     | | || |  | |       |
Db     109144 TAATATATATAATATAATATAACTTATTATATATTATATATTTATATATAAATAAAATAT 109203

Qy        247 GATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAA 306
              |  |       | | ||   |||| | |||     ||| | ||  |  | |||   ||| |
Db     109204 GTACTATATTAATATATGAATATATCTAATATTAATATACAATATATAAATTTATAAATA 109263

Qy        307 AATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGG 366
              |||| | |  |  | | ||||  ||  |  |||  |  | ||  ||   | ||
Db     109264 TATAATGATTATATATTATATAATATATAAAATATATATTATGTAGGGAATCTGAATTTA 109323

Qy        367 AAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAA 426
                ||  ||| ||      || |  | | | | ||  |  |  | ||  || ||| |||
Db     109324 TTTATGTATTTATGTACATATATAAGGTAGGGAATATATATATATGTATTAGGTAGGGAA 109383

Qy        427 AA 428
              |
Db     109384 TA 109385
```

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 13.0%; Score 58.2; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.0012;
Matches 15; Conservative 226; Mismatches 154; Indels 0; Gaps 0;

```
Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              || || ||| |||||  |  |   |     :::::::::: :: : :::  ::::   ::
Db       1461 AGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1402

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
              ::::::  ::: : :::::: ::::::  :: :   :::::::::::: : ::  :  :
Db       1401 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1342

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              ::  :::    :: ::     ::  :  ::  : :   :::::  :: ::::    ::
Db       1341 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1282

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              :   : : :::  :: : :: ::         :    ::::  :  ::: ::   :::
Db       1281 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1222

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              ::: :  : :: : :::     : :: ::: ::::: :::::::  :::::  : :::::
Db       1221 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1162

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
              ::: :  ::::::: :::::: :::  ::::: :: ::  :: : :::  :::::::   :
Db       1161 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1102

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAG 400
              :::::  ::   :  ::: ::   :: ::::::::
Db       1101 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1067
```

RESULT 3
US-08-998-416-186/c
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1074RP
US-08-998-416-186

  Query Match             12.5%;  Score 55.8;  DB 3;  Length 615;
  Best Local Similarity   46.9%;  Pred. No. 0.003;
  Matches  174;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
              |||| |  ||  |||| | |||  || || | ||  | || | | |  | ||
Db        603 ATATTTTATTAATTAATTAAATAATGATAAAATATTAATTAAATAATTAATTTAATTAAA 544

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
              | | ||| | | |  || | |  | ||| ||| | ||  ||  |||
Db        543 TAAATAAATAAATAAAATGAATATAATTTATTAAATAATATAAAAATTAATAATAAGAAA 484

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187
              | | || ||     ||| | | || | | || | |||| ||  |       |
Db        483 TTAAAGTTAAAATTAATTTTAATAATAATTCTTATAAAAAGTTAAATAATATAAATCAAC 424

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247
              ||||  || || |||  | | |  |  |  |||       ||       ||
Db        423 ATAATATTTATAAAAATAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAAT 364

Qy        248 ATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAA 307
              | | |     | | |  |||   ||   ||||| || |  | ||| ||  | ||
Db        363 TAAATCTTTATAATAATAAAATATATTATTTTAATAAACAATTAATAAATAATATATTAA 304

Qy        308 ATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGA 367
              |||| | ||   |  ||||| ||| |  || |  | ||| |||| |  | ||  |
Db        303 TTATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATATCTAATAATATTTTAA 244

Qy        368 AAATTGATTCA 378
              || | ||| |
Db        243 TAACTAATTTA 233


RESULT 4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             12.4%;  Score 55.6;  DB 2;  Length 19124;
  Best Local Similarity   50.8%;  Pred. No. 0.0046;
  Matches  185;  Conservative    0;  Mismatches  174;  Indels    5;  Gaps    2;

Qy          9 TATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGA 68
              |||| | | |  ||  |  | |  | || || |||| ||  ||  |  | || | |
Db      15636 TATAAAAAAAAATAAAAACAAAAGAAGAAAAAAAAAAACATTAAAAAAAAAAAAATATATA 15695

Qy         69 AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATAT---GGTTATCA 125
              | |||   ||  || || | | |||| |  ||  ||||  ||||       | |
Db      15696 TCATAAAAACAAAAAAAAAAGAAAAAAATATATTAAAATAAAAATATATATCATAAAATA 15755

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              ||   | || ||  | | |||| |  |||| || ||  ||| ||| |||  ||
Db      15756 AAAAAAAATTAAAAAAATGTTAAAAAAAAAATATATACATAAAATAAAAAAAATTTATTT 15815

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              |   | | | |  |||| |  ||  | |  |  ||  | |||| | |  |
Db      15816 AAATAAAAAAAAAATAATAAATAAAAAAAATTTAATTAAATAAAAAAAAATAATAAATAAAA 15875

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              | || | ||  | |||   || |  ||  | ||| ||| |  || |  || ||||
Db      15876 AAATTTAATTAAATAAAAAAAAATTAAAAAA--ATTTAATGAAATAAAAAAAAATAAAAA 15933

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
              ||  |   |||  || |||||| ||| | | ||  |  | | |  | | | | |  |
Db      15934 AATTTAATTAAATAAAAAAAATAAAATAAAATTAATTACATGCACATATACATACATATA 15993

Qy        366 GAAA 369
              | ||
Db      15994 TATA 15997


RESULT 5
US-09-801-861-3/c
; Sequence 3, Application US/09801861
; Patent No. 6492154
; GENERAL INFORMATION:
;  APPLICANT:  YAN, Chunhua et al.
;  TITLE OF INVENTION:  ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE:  CL001098
;  CURRENT APPLICATION NUMBER: US/09/801,861
;  CURRENT FILING DATE:  2001-03-09
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 53332
;   TYPE: DNA
;   ORGANISM: Human
US-09-801-861-3

  Query Match             12.2%;  Score 54.6;  DB 4;  Length 53332;
  Best Local Similarity   49.3%;  Pred. No. 0.008;
  Matches  178;  Conservative    0;  Mismatches  174;  Indels    9;  Gaps    1;
```

```
Qy         16 ATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTCAA 75
Db      31625 ATATATAAATATATAAATATATAAAATATATATAATATATAAATATATAAAATATATAAT 31566

Qy         76 ATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTT 135
Db      31565 ATATATAAATATATAAATATATATAAATATATAAATATATATAAATATAAATATATATAA 31506

Qy        136 CAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATA 195
Db      31505 TATATATATAAATATATATAAAATATATATAAATATATAAAAATAT---------AAAAA 31455

Qy        196 GATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATG 255
Db      31454 TATAAATATAATAAATATATAAAAATATATAAATATATAAATATATACAAATATATATAA 31395

Qy        256 GCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAA 315
Db      31394 ATATATAAATATATAAAATATAAATATATAACTATATAAATATATATAAAATATATAAAT 31335

Qy        316 AAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTGAT 375
Db      31334 ATAAATATATATAAAAATATATACGTGATTGCCATTTTTAGAATTATTATTAAAATATAA 31275

Qy        376 T 376
Db      31274 T 31274
```

RESULT 6
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 12.2%; Score 54.4; DB 2; Length 19124;
Best Local Similarity 50.0%; Pred. No. 0.008;
Matches 189; Conservative 0; Mismatches 186; Indels 3; Gaps 2;

```
Qy          9 TATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGA 68
Db      18330 TGTTGATATGTAGAATTACTTTATGTTTTTTGTTTGATAAAAGAATTAAGAAATTAAAAA 18271

Qy         69 AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAAC 128
Db      18270 GGAATAAAGTTAAATAAAAATAAGAAATTAGTTATATTTATAAAAATA--ACAAGAAATA 18213

Qy        129 TGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATC 188
Db      18212 TTATTGTATATTATATTAATTTATTTTAATAGGAAACTATATAATATTGTATTATAAATA 18153

Qy        189 CCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGA 248
Db      18152 ATTTTTATATGAGATTATATATTTTTTTTTCGCTCGGATTATTCAGAATTAGAGTAACAA 18093

Qy        249 TATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAA 308
Db      18092 AAATAAAACACATATATAAAACACATATATAAATACACATTTAAATAATATATTATAAAA 18033

Qy        309 TATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAA 368
Db      18032 TATCACAAATA-AATAAATAAATTAATACTGGAAAAAAATGAAAAAATCGAAAATAGAAG 17974

Qy        369 AATTGATTCATTGGATAA 386
Db      17973 GATATATATATTAGAAAA 17956
```

RESULT 7
US-09-627-122-21/c
; Sequence 21, Application US/09627122
; Patent No. 6472521
; GENERAL INFORMATION:
; APPLICANT: Uhlmann, Eugen
; APPLICANT: Greiner, Beate
; APPLICANT: Unger, Eberhard
; APPLICANT: Gothe, Gislinde
; APPLICANT: Schwerdel, Marc
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN eg5
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: 02481.1678
; CURRENT APPLICATION NUMBER: US/09/627,122
; CURRENT FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-627-122-21

Query Match 12.0%; Score 53.8; DB 4; Length 5340;
Best Local Similarity 47.7%; Pred. No. 0.0093;
Matches 157; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

```
Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       2013 ATATGATGTTTTTGAAAAGGAGGATAATATTATAATTAATAATTATAAAAAAATGATAAA 1954

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       1953 TGAAGATATACAAGAATTTAAAGAAAATATTTTGAAAGATATTGAATTACTTATAAATAA 1894
```

```
Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              || ||    |   | |  ||  ||   | | |   || ||| | ||        ||
Db       1893 AAATGTGGAAGATATTTATGAAAAAATGAATAACAAGTTCGATTTATTAAATAAAGAATT 1834

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              ||   |||       || | ||||    ||    || |||| | | || || | ||
Db       1833 ATATGATAAGTCAAAAAAAAATTATAAATCTGTTTTTATGAATTCTGTAAAAAAGATAGA 1774

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              |||| |||   |||  ||||| |  |||   || |||  ||| || || | |
Db       1773 TGATACTATACAACATTTTTCTGAATGCAATATTAATGATGAAAAGGAATATAATAATAT 1714

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAA 334
              |||| | | |   ||| || | |  ||
Db       1713 TAATAATGATATTAAAAGGAAGATAACAA 1685
```

RESULT 8
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 11.7%; Score 52.4; DB 3; Length 837;
Best Local Similarity 48.1%; Pred. No. 0.015;
Matches 178; Conservative 0; Mismatches 191; Indels 1; Gaps 1;

```
Qy          9 TATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGA 68
              | |||   ||  ||| |  || | |    | |  ||  | |  || ||||    | | | |
Db        601 TTTATTAATTAATTAAATAATGATAAAATATTAATTAAATAATTAATTTAATTAAATAAA 542

Qy         69 AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAAC 128
              ||||| || ||| | | | || | |  || |  |||| | || | | | ||
Db        541 TAAATAAATAAAATGAAATATAATTTATTAAATAATATAAAAATTA-ATAATAAGAAATT 483

Qy        129 TGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATC 188
              || || | ||||||| ||  |  ||| ||  ||||| || |  |  |
Db        482 AAAGTTAAAATTAATTTTAATAATAATTCTTATAAAAAGATTAAATAATATAAATCAACA 423

Qy        189 CCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGA 248
              ||||  || || |||  | | |  | | |  |||        ||        ||
Db        422 TAATATTTATAAAAATAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAATT 363

Qy        249 TATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAA 308
              | | |     | | |  |||   ||  ||||| || |  | || || |  ||
Db        362 AAATCTTTATAATAATAAAATATATTATTTTAATAAACAATTAATAAATAATATATTAAT 303

Qy        309 TATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAA 368
              |||| | ||   |  ||||| ||| | || |  | ||| |||| | | ||  |
Db        302 TATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATATCTAATAATATTTTAAT 243

Qy        369 AATTGATTCA 378
              || | |||  |
Db        242 AACTAATTTA 233
```

RESULT 9
US-08-998-416-1137/c
; Sequence 1137, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1692RP
US-08-998-416-1137

  Query Match             11.5%;  Score 51.2;  DB 3;  Length 636;
  Best Local Similarity   47.8%;  Pred. No. 0.025;
  Matches  178;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
Db        604 ATATTTTATTAATTAATTAAATAATGATAAAATATTAATTAAATAATTAATTTAATTAAA 545

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
Db        544 TAAATAAATAAATAAAATGAATATAATTTATTAAATAATATAAAAATTAATAATAAGAAA 485

Qy        128 CTGGA-TTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTA 186
Db        484 TTAAAGTTAAAATTAATTTTAATAATAATTCTTATAAAAAGATTAAATAATATAAATCAA 425

Qy        187 TCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAA 246
Db        424 CATAATATTTATAAAAATAGATATTATAATAAAAATAATAATTTACAATATTTAAATAAA 365

Qy        247 GATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAA 306
Db        364 TTAAATCTTTATAATAATAAAATATATTATTTTAATAAACAATTAATAAATAATATATTA 305

Qy        307 AATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGG 366
Db        304 ATTATTGATAATCTATTTAATAATTTATTAAAGAAAATAATAATATCTAATAATATTTTA 245

Qy        367 AAAATTGATTCA 378
Db        244 ATAACTAATTTA 233


RESULT 10
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
;  APPLICANT:  Inselburg, J. et al.
;  TITLE OF INVENTION:  GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
;  TITLE OF INVENTION:  AND USES THEREFOR
;  FILE REFERENCE:  JII-002CNCP
;  CURRENT APPLICATION NUMBER: US/08/213,419B
;  CURRENT FILING DATE:  1994-03-14
;  PRIOR APPLICATION NUMBER:  US 07/870,506
;  PRIOR FILING DATE:  1992-04-17
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6124
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2407)..(2439)
;   NAME/KEY: CDS
;   LOCATION: (2598)..(3404)
;   NAME/KEY: CDS
;   LOCATION: (3580)..(3720)
;   NAME/KEY: CDS
;   LOCATION: (3850)..(5835)
US-08-213-419B-3

  Query Match             11.1%;  Score 49.8;  DB 4;  Length 6124;
  Best Local Similarity   48.4%;  Pred. No. 0.06;
  Matches  138;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         65 AAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATC 124
Db       2443 AAGAATGAAAAAAAAAAAAAAAAGGAAAAGGAAAAGAAAAAGGAAAAGAAAAAGGAAAAG 2502

Qy        125 AAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGT 184
Db       2503 AAAAAGAAAAACAAATATGTAAAAATATAATTATTATATAATAAATAATATAATATTTTT 2562

Qy        185 TATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGG 244
Db       2563 ACGCATACACAAACATTTGTCATTATTTTTTTTTAGGTGTTATATTTAACAAAAATGTTA 2622

Qy        245 AAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAA 304
Db       2623 TAAAATGTACAGGAGAAAGTCAAACAGGTAATACAGGAGGAGGTCAAGCAGGTAATACAG 2682

Qy        305 AAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTA 349
Db       2683 TAGGAGATCAAGCAGGTAGTACAGGAGGAAGTCCACAAGGTAGTA 2727

RESULT 11
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoc
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 11.0%; Score 49; DB 4; Length 1664976;
Best Local Similarity 45.5%; Pred. No. 0.15;
Matches 175; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

```
Qy         36 AGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTCAAATACAAGGAACAAAAGATCA 95
              |||| ||||| || | || || | |||||  | ||| | ||     | | |
Db     844063 AGAAAAAGAAAAACCTAAAGAAGTAAAAAAAAGAGATTAAAAAAGAAACTGAAGAGAAAAT 844004

Qy         96 ATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTTCAACGATTTTTAATCATTGG 155
              | |  |||| || ||||| || | |  ||   | |    ||| |||| |||  |
Db     844003 AGAAAAAGAAAAAATAGAATTTGTTAAGAAAGAAGAAAAAGAACAATTTATAAAAAAATC 843944

Qy        156 TTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATAGATCAATATAATAACTCTTC 215
              | ||   | |   | |||||||      |     |    | || | || ||
Db     843943 TGATGAAGATGTTGAAGAGAAATTAAAACAACTAATTTCCAAAGAAGAGAAAAAAGAAGA 843884

Qy        216 TTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATGGCACAAATTTCTATAACGAA 275
```

```
Db      843883 TTTCGATGCTGAGAGGGCTAAAAGATATGAACACATAACAAAAATAAAAGAGAGTGTAAA 843824

Qy         276 TGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAAAAGGCAAGAAATAAATTAAG 335

Db      843823 TAGTAGAATAAAATGGATAGCTAAAGACATCGATGCCGTGATTGAGATATATGAAGATTC 843764

Qy         336 AGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTGATTCATTGGATAATTTGGTGAA 395

Db      843763 AGACGTGTCTGGAAAATCCACATGCACTGGAACTATTGAGGACTTCGTTAAATACTTTAG 843704

Qy         396 AGAAGTAGGTGGAACTTTACAAGTT 420

Db      843703 AGACAGATTTGAAAGATTAAAGGTT 843679


RESULT 12
US-09-601-198-70
; Sequence 70, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
;  APPLICANT: Cassell, Gail H.
;  APPLICANT:  Chen, Ellson Y.
;  APPLICANT:  Glass, Jennifer S.
;  APPLICANT:  Glass, John I.
;  APPLICANT:  Heiner, Cheryl R.
;  APPLICANT:  Lefkowitz, Elliot
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
;  TITLE OF INVENTION:  UREALYTICUM
;  FILE REFERENCE: UAB-13452/22
;  CURRENT APPLICATION NUMBER: US/09/601,198
;  CURRENT FILING DATE:  2000-12-08
;  PRIOR APPLICATION NUMBER: 60/073,189
;  PRIOR FILING DATE: 1998-01-30
;  NUMBER OF SEQ ID NOS: 181
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 2427
;   TYPE: DNA
;   ORGANISM: Ureaplasma urealyticum
US-09-601-198-70

  Query Match            10.8%;  Score 48.2;  DB 4;  Length 2427;
  Best Local Similarity  47.9%;  Pred. No. 0.11;
  Matches  171;  Conservative    0;  Mismatches  183;  Indels    3;  Gaps    1;

Qy          72 TCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGG 131

Db        1701 TGAAAAAGATCCAGATACTGATAAAGATTTACTAGTTGAAGAGCGTGATTATACTCCTTG 1760

Qy         132 ATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCA 191

Db        1761 AATTAATACATCACTAATTAATATATATCAACTTCCACAATTTTTTAATATCTTATTTAA 1820

Qy         192 TATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATAT 251

Db        1821 TCAATCTGAATGATTATTATCAATTCAATATCTAAAAGAAATGAGTGATTATAAAAAATA 1880

Qy         252 TATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATAT 311

Db        1881 TTTACTACAAATTTTTGAACCATATCAAGATAAAGAATGGATTTTTAATTTAAAACA--- 1937

Qy         312 TAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAAT 371

Db        1938 TAAAGATTGACAAGATGGCGAACAAGTGGTTGCTATTTATAAAAATGCAGATAGTAATAG 1997

Qy         372 TGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAAAA 428

Db        1998 CGACCAAAAAGAATATTATTTTAACGGCTATGTTAAAAATCAACGAGCTTATATTAA 2054


RESULT 13
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-855A-1

  Query Match            10.7%;  Score 48;  DB 2;  Length 8920;
  Best Local Similarity  46.0%;  Pred. No. 0.14;
  Matches  198;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    1;

Qy           8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67

Db         877 AGATATATATACTTTTAATAACAACTTTGTGATGTTAAAAGAATAAAACTGTTTAAGACC 818

Qy          68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127

Db         817 TATGATTCAGAGAATATCCCCAATAATTATATATATATATATTTATATATATCTAT 758

Qy         128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187

Db         757 ATTATTTTTTCCCATTTTTCTTTTTTTTTTTTATACATTTAT--ATTTATGTTTAAAATA 700

Qy         188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247

Db         699 TTTATAAATTTACATATACAAGTTCATTTTTCATATGTAAATTTTTTTTTTTTCTTTTTT 640

Qy         248 ATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAA 307

Db         639 TTTTTTTTTTTTTTTTTTTTTTAAATTAGTAGAATTACTATTTTATAAACTAAGAAAAA 580

Qy         308 ATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGA 367

Db         579 AAATAAATAAATGAATAAATTAATAAATAAATATATAAATAAAATTATAGGAACCACAAT 520

Qy         368 AAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAAA 427

Db         519 ATTGGGGAGTATTATATATTGTGTATAATATATAGGATGGTTTTATTATAAGAAGTGTAA 460
```

```
Qy        428 ACCCCGATGA 437

Db        459 AACTATATTA 450


RESULT 14
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 8920
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-150-741-1

  Query Match            10.7%;  Score 48;  DB 3;  Length 8920;
  Best Local Similarity  46.0%;  Pred. No. 0.14;
  Matches  198;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    1;

Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67

Db        877 AGATATATATACTTTTAATAACAACTTTGTGATGTTAAAAGAATAAAACTGTTTAAGACC 818

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127

Db        817 TATGATTCAGAGAATATCCCCAATAATTATATATATATATATATTTATATATATATCTAT 758

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187

Db        757 ATTATTTTTTCCCATTTTTCTTTTTTTTTTTATACATTTAT--ATTTATGTTTAAAATA 700

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247

Db        699 TTTATAAATTTACATATACAAGTTCATTTTTCATATGTAAATTTTTTTTTTTTCTTTTTT 640

Qy        248 ATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAA 307

Db        639 TTTTTTTTTTTTTTTTTTTTTTTTAAATTAGTAGAATTACTATTTTATAAACTAAGAAAAA 580

Qy        308 ATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGA 367

Db        579 AAATAAATAAATGAATAAATTAATAAATAAATATATAAATAAAATTATAGGAACCACAAT 520

Qy        368 AAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAAA 427

Db        519 ATTGGGGAGTATTATATATTGTGTATAATATATAGGATGGTTTTATTATAAGAAGTGTAA 460

Qy        428 ACCCCGATGA 437

Db        459 AACTATATTA 450


RESULT 15
US-09-601-198-75/c
; Sequence 75, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
;  APPLICANT: Cassell, Gail H.
;  APPLICANT:  Chen, Ellson Y.
;  APPLICANT:  Glass, Jennifer S.
;  APPLICANT:  Glass, John I.
;  APPLICANT:  Heiner, Cheryl R.
;  APPLICANT:  Lefkowitz, Elliot
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
;  TITLE OF INVENTION:  UREALYTICUM
;  FILE REFERENCE: UAB-13452/22
;  CURRENT APPLICATION NUMBER: US/09/601,198
;  CURRENT FILING DATE:  2000-12-08
;  PRIOR APPLICATION NUMBER: 60/073,189
;  PRIOR FILING DATE: 1998-01-30
;  NUMBER OF SEQ ID NOS: 181
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
;   LENGTH: 5652
;   TYPE: DNA
;   ORGANISM: Ureaplasma urealyticum
US-09-601-198-75

  Query Match            10.6%;  Score 47.2;  DB 4;  Length 5652;
  Best Local Similarity  47.1%;  Pred. No. 0.2;
  Matches  145;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy         91 GATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGATTTCAACGATTTTTAATC 150

Db       4542 GATCTTGAATCAATTGGTAAATCAAGTGGTAATCAAAATTATTTACTAATTGATGCAACC 4483

Qy        151 ATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATATAGATCAATATAATAAC 210

Db       4482 ATTGGTGGCCAACAACAAAAAAATTAAATACTACATGTTTAAGGATGAATTCACTACTAAA 4423

Qy        211 TCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTATGGCACAAATTTCTATA 270

Db       4422 TCAAGTTCAAATGATGAATTTATTCAAAATGATAATAGTATTTAGCATCAAGTTTACGT 4363

Qy        271 ACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTAAAAAGGCAAGAAATAAA 330

Db       4362 GAAATAATTACTAAAAATTTTAATGATTATATTGATCAAAAACAAAATTTTAAAAATGTT 4303

Qy        331 TTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTGATTCATTGGATAATTTG 390

Db       4302 TATGATGTTGAATCTACATTTAAAAATAAACAATTTTATATTTATGATATTGATGGTAAT 4243

Qy        391 GTGAAAGA 398

Db       4242 GGTAAAAA 4235


Search completed: February 16, 2004, 12:52:52
Job time : 72 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 06:13:40 ; Search time 277 Seconds
(without alignments)
4356.133 Million cell updates/sec

Title: US-09-980-054A-11
Perfect score: 447
Sequence: 1 atgtcagatatagatataga.....accccgatgatatgtggtga 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		447	100.0	447	22	AAC86750	DNA encoding a Can
2		442.2	98.9	447	22	AAS23431	Candida albicans e
3		442.2	98.9	447	24	ABZ31775	Candida albicans e
4		324.4	72.6	326	22	AAC86756	Probe for DNA enco
c	5	70.4	15.7	4985	24	ABQ75107	Anopheles gambiae
c	6	65	14.5	65	24	ABZ28869	Candida gene relat
c	7	62.6	14.0	9539	22	AAS45346	Chemically pretrea
c	8	62.6	14.0	9539	24	ABK28179	DNA transcription
	9	61.2	13.7	168575	22	AAH21613	Human hypocretin r
c	10	60.2	13.5	17183	24	ABL32487	Human immune syste
c	11	59.8	13.4	6109	24	ABL32326	Human immune syste
c	12	59.8	13.4	6109	24	AAS61077	Human gene regulat
	13	59.8	13.4	9789	17	AAT41852	cDNA encoding Plas
c	14	59.8	13.4	12237	24	ABL34358	Human immune syste
c	15	59.4	13.3	446	25	ABX38235	Bovine EST associa
	16	59.2	13.2	11143	23	ABL12834	Drosophila melanog
c	17	59.2	13.2	12142	24	ABL33673	Human immune syste
c	18	57.8	12.9	516	25	ABX40620	Bovine EST associa
c	19	57.6	12.9	9539	22	AAS45347	Chemically pretrea
c	20	57.6	12.9	9539	24	ABK28180	DNA transcription
c	21	57.4	12.8	6641	24	ABL54335	Chemically treated
c	22	57.4	12.8	6641	24	ABL32314	Human immune syste
	23	57.4	12.8	32392	24	ABL56203	AmEPV genome fragm
	24	57.2	12.8	4311	21	AAA70133	Plasmodium falcipa
	25	57	12.8	5940	21	AAA70105	Plasmodium falcipa
c	26	56.8	12.7	17131	24	ABL33053	Human immune syste
c	27	56.6	12.7	6763	24	ABQ67067	Human angiogenesis
c	28	56.4	12.6	402	22	AAI82045	Human polynucleoti
	29	56.4	12.6	1686	16	AAQ87587	DNA encoding Leuco
c	30	56.4	12.6	6301	24	ABL32053	Human immune syste
c	31	55.8	12.5	6286	22	AAS46591	Tumour suppressor
	32	55.6	12.4	1407	24	ABL56226	AmEPV serine threo
c	33	55.6	12.4	9963	24	ABL32694	Human immune syste
	34	55.6	12.4	19124	18	AAT72882	Plasmodium var-7 g
	35	55.6	12.4	19124	21	AAZ98287	Plasmodium var-7 p
	36	55.6	12.4	50000	24	ABL56201	AmEPV genome fragm
	37	55.2	12.3	1998	21	AAA70212	Plasmodium falcipa
c	38	55.2	12.3	4590	7	AAN60472	Sequence encoding
c	39	55.2	12.3	6175	24	ABL33307	Human immune syste
	40	55.2	12.3	6621	21	AAA70188	Plasmodium falcipa
c	41	55	12.3	55	24	ABZ28979	Candida gene relat
c	42	55	12.3	80	22	AAS23676	Tetracycline promo
c	43	54.8	12.3	419	25	ABX46069	Bovine EST associa
c	44	54.8	12.3	509	24	ABQ16728	Oligonucleotide fo
	45	54.8	12.3	509	24	ABQ16729	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAC86750
ID AAC86750 standard; DNA; 447 BP.
XX
AC AAC86750;
XX
DT 02-APR-2001 (first entry)
XX
DE DNA encoding a Candida albicans protein CaNL260.
XX
KW CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CaNL260; CaDR361;
KW antifungal; fungal infection; pathogenic fungi; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 1..447
FT /*tag= a
XX
PN WO200075305-A2.
XX
PD 14-DEC-2000.
XX
PF 08-JUN-2000; 2000WO-FR01567.
XX
PR 09-JUN-1999; 99FR-0007250.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Lalanne J, Rocher C;
XX

DR WPI; 2001-050024/06.
DR P-PSDB; AAB30859.
XX
PT New polynucleotides from Candida albicans and their derived proteins,
PT useful for diagnosis and treatment of fungal infections and for drug
PT screening -
XX
PS Claim 4; Page 77-78; 89pp; French.
XX
CC The present sequence encodes a Candida albicans protein. The
CC specification describes genes CaDRA472, CaDR489, 1CaDR527, 2CaDR527,
CC CaFLO24, CaNL260, and CaDR361. These genes are essential for
CC survival, and so are good targets for antifungal agents. The Candida
CC albicans genes and their derived proteins are used to screen compounds
CC for the ability to inhibit the activity of the protein, i.e. for
CC antifungal activity. The proteins are also used to generate a protective
CC antibody response against fungal infections in mammals. The Candida
CC albicans proteins and genes, and their antibodies, are used for
CC diagnosing fungal infections, specifically C. albicans (in standard
CC amplification, hybridisation or immunological assays, and for studying
CC pathogenic fungi.
XX
SQ Sequence 447 BP; 192 A; 38 C; 82 G; 135 T; 0 other;

Query Match 100.0%; Score 447; DB 22; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.2e-70;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447
```

RESULT 2
AAS23431
ID AAS23431 standard; DNA; 447 BP.
XX
AC AAS23431;
XX
DT 04-DEC-2001 (first entry)
XX
DE Candida albicans essential gene CaYNL260C.
XX
KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection; ds.
XX
OS Candida albicans.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05551.
XX
PR 18-FEB-2000; 2000US-0183534.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
DR P-PSDB; AAU15103.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes -
XX
PS Claim 22; Page 167; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAS23381-AAS23442 represent C. albicans essential genes.
XX
SQ Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Query Match 98.9%; Score 442.2; DB 22; Length 447;
Best Local Similarity 99.3%; Pred. No. 8.7e-70;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              |||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447
```

```
RESULT 3
ABZ31775
ID   ABZ31775 standard; DNA; 447 BP.
XX
AC   ABZ31775;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential gene SEQ ID NO 6062.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US49486.
XX
PR   29-DEC-2000; 2000US-259128P.
PR   20-FEB-2001; 2001US-0792024.
PR   22-AUG-2001; 2001US-314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   P-PSDB; ABP73225.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele
PT   of a gene and placing other allele of the gene under conditional
PT   expression -
XX
PS   Claim 37; SEQ ID NO 6062; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans gene used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office.
XX
SQ   Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

  Query Match             98.9%;  Score 442.2;  DB 24;  Length 447;
  Best Local Similarity   99.3%;  Pred. No. 8.7e-70;
  Matches  444;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGGAAAATATTTTGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ACTTGGAAAATTGATTCATTAGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Qy        421 AGTGAAAACCCCGATGATATGTGGTGA 447
              |||||||||||||||||||||||||||
Db        421 AGTGAAAACCCCGATGATATGTGGTGA 447


RESULT 4
AAC86756
ID   AAC86756 standard; DNA; 326 BP.
XX
AC   AAC86756;
XX
DT   02-APR-2001  (first entry)
XX
DE   Probe for DNA encoding a Candida albicans protein CaNL260.
XX
KW   CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CaNL260; CaDR361;
KW   antifungal; fungal infection; pathogenic fungi; probe; ss.
XX
OS   Candida albicans.
XX
PN   WO200075305-A2.
XX
PD   14-DEC-2000.
XX
PF   08-JUN-2000; 2000WO-FR01567.
XX
PR   09-JUN-1999;   99FR-0007250.
XX
PA   (HMRI ) HOECHST MARION ROUSSEL.
XX
PI   Lalanne J,  Rocher C;
XX
DR   WPI; 2001-050024/06.
XX
PT   New polynucleotides from Candida albicans and their derived proteins,
PT   useful for diagnosis and treatment of fungal infections and for drug
PT   screening -
XX
PS   Example 5; Page 85-86; 89pp; French.
XX
CC   The present sequence represents a probe for DNA encoding a Candida
CC   albicans protein. The specification describes genes CaDRA472, CaDR489,
CC   1CaDR527, 2CaDR527, CaFLO24, CaNL260, and CaDR361. These genes are
CC   essential for survival, and so are good targets for antifungal agents.
CC   The Candida albicans genes and their derived proteins are used to screen
```

```
CC   compounds for the ability to inhibit the activity of the protein,
CC   i.e. for antifungal activity. The proteins are also used to generate a
CC   protective antibody response against fungal infections in mammals. The
CC   Candida albicans proteins and genes, and their antibodies, are used for .
CC   diagnosing fungal infections, specifically C. albicans (in standard
CC   amplification, hybridisation or immunological assays, and for studying
CC   pathogenic fungi.
XX
SQ   Sequence 326 BP; 146 A; 27 C; 53 G; 100 T; 0 other;

  Query Match             72.6%;  Score 324.4;  DB 22;  Length 326;
  Best Local Similarity   99.7%;  Pred. No. 6.1e-49;
  Matches  325;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         12 AGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGG 71
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 AGATATAGATAATGTATTAAATTTAGAAGAAGATCAATATGAATTAGGATTTAAAGAAGG 60

Qy         72 TCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGG 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGG 120

Qy        132 ATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCA 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCA 180

Qy        192 TATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATAT 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATAT 240

Qy        252 TATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATAT 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATAT 300

Qy        312 TAAAAAGGCAAGAAATAAATTAAGAG 337
              |||||||||||||||||||||||||
Db        301 TAAAAAGGCAAGAAATAAATTAAGAG 326


RESULT 5
ABQ75107/c
ID   ABQ75107 standard; cDNA; 4985 BP.
XX
AC   ABQ75107;
XX
DT   01-NOV-2002  (first entry)
XX
DE   Anopheles gambiae odourant receptor 2 genomic DNA SEQ ID NO:10.
XX
KW   Anopheles gambiae; mosquito; olfactory gene; arrestin 1; pest control;
KW   odourant receptor; olfaction; gene; ds.
XX
OS   Anopheles gambiae.
XX
PN   WO200259274-A2.
XX
PD   01-AUG-2002.
XX
PF   28-JAN-2002; 2002WO-US02549.
XX
PR   26-JAN-2001; 2001US-264649P.
PR   24-JAN-2002; 2002US-0056405.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Zwiebel LJ;
XX
DR   WPI; 2002-627421/67.
DR   P-PSDB; ABP52835.
XX
PT   New mosquito olfaction polypeptides and polynucleotides, useful for
PT   mosquito management, i.e. controlling the pest and disease vectors, or
PT   for identifying pest control agents -
XX
PS   Disclosure; Fig 4a; 96pp; English.
XX
CC   The present invention describes a purified Anopheles gambiae olfaction
CC   polypeptide comprising a 383, 394, 380, 411, 412, 391, 157 or 401 residue
CC   amino acid sequence (see ABP52833 to ABP52840) (S1), a conservatively
CC   modified amino acid sequence of them, or a sequence of (S1) with at least
CC   20 consecutive residues. Also described: (1) an isolated polynucleotide
CC   comprising: (a) a nucleotide sequence encoding the purified Anopheles
CC   gambiae olfaction polypeptide; or (b) a nucleotide sequence that
CC   hybridises under stringent conditions to a hybridisation probe comprising
CC   a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence
CC   (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (S2), or its
CC   complement; and (2) a method for identifying an agent that binds to
CC   mosquito olfaction molecules comprising: (a) providing an isolated
CC   mosquito olfaction molecule; (b) contacting a test agent with the
CC   isolated mosquito olfaction molecule; and (c) detecting specific binding
CC   of the test agent to the isolated mosquito olfaction molecule, where the
CC   presence of specific binding identifies the test agent as a mosquito
CC   olfaction-binding compound. The mosquito olfaction molecules are useful
CC   for mosquito management, i.e. controlling this pest and disease vector.
CC   A method from the present invention of screening for substances that
CC   modulate arrestin-odourant receptor interaction is useful for identifying
CC   pest control agents. The present sequence represents Anopheles gambiae
CC   odourant receptor 2 genomic DNA from the present invention.
CC   N.B. The features given in figure 4a are tentative and do not directly
CC   encode SEQ ID NO:6 (ABP52835).
XX
SQ   Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 other;

  Query Match             15.7%;  Score 70.4;  DB 24;  Length 4985;
  Best Local Similarity   51.0%;  Pred. No. 0.00057;
  Matches  196;  Conservative    0;  Mismatches  181;  Indels    7;  Gaps    1;

Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              | | ||  | |   | || |   || | | | | || || || |||| |||| |  ||
Db       2850 TAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAATAAGAATAATAATAA 2791

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
              || ||    ||  | ||  |  | || |  |||| | | ||  |||  | |||  | ||
Db       2790 TAATAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTATTATTATGAT 2731

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
              | | || ||   || || |   |||         ||| | |  ||| |    || ||
Db       2730 AATAATGTATAATAATGAATAACAATA-------ATAATAATTAATAATAATAAATTAAT 2678

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              | ||      ||| || |  |||||| | || ||    |  ||| ||   ||||||
Db       2677 TAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAATAATAATAATAATAAT 2618

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              |  ||| ||     || | || ||  |  |||    ||||  |   | | || || |
Db       2617 AATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATAATAATAAT 2558

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              || |||| ||| ||    |  ||||||   |          || ||  | |||  | | |
Db       2557 AATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATAATAAT 2498

Qy        364 TGGAAAATTGATTCATTGGATAAT 387
              || | | ||    |  |||||
Db       2497 AATAATAATAATAATATTAATAAT 2474


RESULT 6
ABZ28869/c
ID   ABZ28869 standard; DNA; 65 BP.
XX
AC   ABZ28869;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida gene related tetracyclin promoter PCR primer SEQ ID NO 2884.
XX
```

KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US49486.
XX
PR 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression -
XX
PS Claim 76; SEQ ID NO 2884; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of a PCR
CC primer used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 65 BP; 20 A; 11 C; 2 G; 32 T; 0 other;

Query Match 14.5%; Score 65; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 65 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 6

Qy 64 AAAGA 68
|||||
Db 5 AAAGA 1

RESULT 7
AAS45346/c
ID AAS45346 standard; DNA; 9539 BP.
XX
AC AAS45346;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated genomic DNA associated with cell cycle #26.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02945.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
PS Claim 1; SEQ ID No 51; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
SQ Sequence 9539 BP; 2665 A; 121 C; 2515 G; 4238 T; 0 other;

Query Match 14.0%; Score 62.6; DB 22; Length 9539;
Best Local Similarity 47.8%; Pred. No. 0.014;
Matches 182; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

Qy 4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
||| | | | || | | || |||||||| || || | ||||| || | | |
Db 6192 TCAAAAAAATTTAAAATATAACCATAAAATTTATAAAAAAACAAATATAAACGAAATTAT 6133

Qy 64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
||| | | || | || || | || |||| | |||| | || | |
Db 6132 AAAACAAAAACGAACTTAAAAATTCAAAAACACATTTTAAACTCAAAAATAAATTATAAA 6073

Qy 124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
| | || | || | | || | | | | |||| ||| || || |
Db 6072 AACAAACGAAATTAAAAAATCGAAACGAAATTTAAAAATCAAAAATAAAATTAAAAAATA 6013

```
Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              |||   |   | ||   | ||| |                 | |||       |   |
Db       6012 ATATAAAAATCAAATTCTAAAAATCATAAAAAAATCTAAAAAATTCAAAAATTTTTCTAA 5953

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              || | | ||      ||| || | ||   || | | ||| ||  |  |  | || ||
Db       5952 AAAAAAAATAAAATTTAAAAATCAAAAATCTATTTAAACAAACAAAATATAACTAATAAA 5893

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              ||||| |||||||  ||| | | || | ||     | |      || |||||| ||| |
Db       5892 AAAAAATTTAAAAAACCAAAATAAAATTAAAATAAAAAAAAACTTAAAACTAAAAAAATT 5833

Qy        364 TGGAAAATTGATTCATTGGAT 384
              ||||   ||  | | ||
Db       5832 CAAAAAAACAATAAAATATAT 5812
```

RESULT 8
ABK28179/c
ID ABK28179 standard; DNA; 9539 BP.
XX
AC ABK28179;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated genomic DNA #27.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer -
XX
PS Claim 1; SEQ ID No 53; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 9539 BP; 2665 A; 121 C; 2515 G; 4238 T; 0 other;

Query Match 14.0%; Score 62.6; DB 24; Length 9539;
Best Local Similarity 47.8%; Pred. No. 0.014;
Matches 182; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

```
Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
              ||| | | |  || |    |  || |||||||| || || |  |||| ||  |  | |
Db       6192 TCAAAAAAATTTAAAATATAACCATAAAATTTATAAAAAAACAAATATAAACGAAATTAT 6133

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
              |||  |      |   |  ||   || | ||   |||| |   ||||  | ||  | |
Db       6132 AAAACAAAAACGAACTTAAAAATTCAAAAACACATTTTAAACTCAAAAATAAATTATAAA 6073

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
               | |   ||  | ||  | |   ||  |  | | |||| ||| ||  ||
Db       6072 AACAAACGAAATTAAAAAATCGAAACGAAATTTAAAAATCAAAAATAAAATTAAAAAATA 6013

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
              |||   |   | ||   | ||| |                 | |||       |   |
Db       6012 ATATAAAAATCAAATTCTAAAAATCATAAAAAAATCTAAAAAATTCAAAAATTTTTCTAA 5953

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
              || | | ||      ||| || | ||   || | | ||| ||  |  |  | || ||
Db       5952 AAAAAAAATAAAATTTAAAAATCAAAAATCTATTTAAACAAACAAAATATAACTAATAAA 5893

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363
              ||||| |||||||  ||| | | || | ||     | |      || |||||| ||| |
Db       5892 AAAAAATTTAAAAAACCAAAATAAAATTAAAATAAAAAAAAACTTAAAACTAAAAAAATT 5833

Qy        364 TGGAAAATTGATTCATTGGAT 384
              ||||   ||  | | ||
Db       5832 CAAAAAAACAATAAAATATAT 5812
```

RESULT 9
AAH21613
ID AAH21613 standard; DNA; 168575 BP.
XX
AC AAH21613;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human hypocretin receptor 2 (HCRTR2) gene SEQ ID NO:1.
XX
KW Human; narcolepsy; hypocretin receptor 2; orexin receptor 2; HCRTR2;
KW diagnosis; ds.
XX
OS Homo sapiens.
XX
PN WO200130991-A2.
XX
PD 03-MAY-2001.
XX
PF 22-AUG-2000; 2000WO-US23021.
XX
PR 25-OCT-1999; 99US-0426290.
XX

PA (DECO-) DECODE GENETICS EHF.
XX
PI Olafsdottir BR, Gulcher J;
XX
DR WPI; 2001-300504/31.
DR P-PSDB; AAB98007.
XX
PT Gene for hypocretin (orexin) receptor 2 (HCRTR2) which is associated
PT with narcolepsy, useful in methods of diagnosis of narcolepsy and
PT pharmaceutical compositions for therapy -
XX
PS Claim 1; Fig 1; 85pp; English.
XX
CC The present invention describes the human hypocretin (orexin) receptor 2
CC (HCRTR2) gene (given in AAH21613), which is associated with narcolepsy.
CC Identification of the HCRTR2 nucleic acid molecule permits the diagnosis
CC of narcolepsy. A method from the present invention is provided for
CC treating narcolepsy by administering to the individual an isolated
CC HCRTR2 nucleic acid in a therapeutically effective amount so that the
CC cells produce native HCRTR2 receptor. The diagnosis of narcolepsy has
CC been difficult to differentiate from other conditions such as chronic
CC fatigue syndrome or other sleep disorders but detection of HCRTR2
CC nucleic acid makes it possible to accurately diagnose narcolepsy.
CC AAH21541 to AAH21612 represent primers used in the identification of the
CC narcolepsy gene in an example from the present invention. AAH21613
CC represents the HCRTR2 gene which encodes the HCRTR2 protein given in
CC AAB98007.
XX
SQ Sequence 168575 BP; 55308 A; 29672 C; 29838 G; 53757 T; 0 other;

Query Match 13.7%; Score 61.2; DB 22; Length 168575;
Best Local Similarity 48.1%; Pred. No. 0.025;
Matches 203; Conservative 0; Mismatches 218; Indels 1; Gaps 1;

```
Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATAT-GAATTAGGATTTAAA 66
              ||||| | ||| ||||||||||||| | ||| |   |  |||| || ||     | |
Db     108964 ATATATAAATATATAATGTATTATAGTTATATATAATATTATATATAAATATATAATGTA 109023

Qy         67 GAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAA 126
              |  |   ||| |     |  | | || ||||| || | | |       |  |  | |
Db     109024 TTATATGTTATATATAATATTACATATAAATATCTATTAATATATATTATTTATATTATA 109083

Qy        127 ACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTA 186
              ||  || |  |  || |   |  |     ||||||  |  || | | | || || ; ||
Db     109084 TCTAAATATATAATATATAACTTATTATTATATATTATAATATAACTTATATATTATATA 109143

Qy        187 TCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAA 246
              |   ||||| |  |  ||| | | || | | |   | | || |  | |       |
Db     109144 TAATATATATAATATAATATAACTTATTATATATTATATATTTATATATAAATAAAATAT 109203

Qy        247 GATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAA 306
              |  |       | | ||   |||| | |||    ||| | ||  |  | |||   ||| |
Db     109204 GTACTATATTAATATATGAATATATCTAATATTAATATACAATATATAAATTTATAAATA 109263

Qy        307 AATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGG 366
              ||| | |  |   |  | |||| || |  |||  | | ||   ||    | ||
Db     109264 TATAATGATTATATATTATATAATATAAAATATATATTATGTAGGGAATCTGAATTTA 109323

Qy        367 AAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAA 426
              ||  ||| ||       || | | | | | ||  | | ||  || ||| |||
Db     109324 TTTATGTATTTATGTACATATATAAGGTAGGGAATATATATATATGTATTAGGTAGGGAA 109383

Qy        427 AA 428
              ||
Db     109384 TA 109385
```

RESULT 10
ABL32487/c
ID ABL32487 standard; DNA; 17183 BP.
XX
AC ABL32487;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 460.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 460; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 17183 BP; 5342 A; 149 C; 3418 G; 8274 T; 0 other;

Query Match 13.5%; Score 60.2; DB 24; Length 17183;
Best Local Similarity 49.9%; Pred. No. 0.037;
Matches 185; Conservative 0; Mismatches 178; Indels 8; Gaps 1;

```
Qy          8 ATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAG 67
              ||| | | ||| || ||  ||    |  || || ||  | || |  |    | ||||
Db       4235 ATAAAAAAATAAATCATAAATACTACCTAAAAAAATTAAAAAACTATAAACCCTCTAAAC 4176

Qy         68 AAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAA 127
              ||    ||  |||   |  |||| || |||||  | ||  | || |         |  ||
Db       4175 AAAAAAAACAACATTTACAAAAACATAAATATACAAAAACATAATATCTCTAAAAAAAAT 4116

Qy        128 CTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTAT 187
              |   |||   |  || |  | |  |  | | | |  || |||||||  ||||      |
Db       4115 CAATATTATTAAAATATAAACTACATAATAAAAAATAATAATTAATTAAATAAATTAAAA 4056

Qy        188 CCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAG 247
              |  | || |  || |||   ||  || ||| ||     |      | ||| ||   ||
Db       4055 CTTACTTATAAAAAAATATATACATTTTTTCCCTATAAACAATAATAAATCATAATAAAT 3996

Qy        248 ATATTATGGCACAAATTTCTATAAC--------GAATGGAGATAAAGAAGTTGAAGATTA 299
              | | |    ||||| |  |||||          | |  | | ||| || |  |  || |
Db       3995 TTTTAAATATACAAAATAACATAACCACATTTATATTTTAAAAAAATAAATATACAATAA 3936

Qy        300 TGAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGA 359
              |  |||||||||   ||    || ||| | | | | |   |||  |  ||| | | || |
```

```
Db       3935 TATAAAAAATATACCAACCCAAAAAAAAACAATTAAAAAAATAAATTATATCATTCAATA 3876

Qy        360 AACTTGGAAAA 370
              | |    ||||
Db       3875 ATCCAATAAAA 3865
```

RESULT 11
ABL32326/c
ID ABL32326 standard; DNA; 6109 BP.
XX
AC ABL32326;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 299.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 299; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6109 BP; 1232 A; 195 C; 1734 G; 2944 T; 4 other;

  Query Match   13.4%; Score 59.8; DB 24; Length 6109;
  Best Local Similarity   46.3%; Pred. No. 0.043;
  Matches 196; Conservative   0; Mismatches 227; Indels   0; Gaps   0;

```
Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              |||| | | ||| ||||  | ||||| | ||  | |  ||||   |     ||  |
Db       2081 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 2022

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
              | |      ||||| |   |  || | || ||||  || | | ||||  | ||   ||
Db       2021 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAACGAAAAAATAAATAAATAAAT 1962

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
```

```
                 || |  ||     |       | ||| | |   || ||  |  ||  ||| ||    |   |
Db       1961 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAAATAAATAAATAAAT 1902

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              |   |  |   |  |  |||    |  | | |    ||| | |  |  |||     |
Db       1901 AAATAAATAAATAAATAAATATATAAATAAATAAATAATAAATAAATATATAAATAAATA 1842

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              |   | ||    |  || |      |  ||| | | ||| |     || |     | ||
Db       1841 AATAAATAAATAAATAAATATATAAATAAATAAATAAAAAAATAAAAAAATAAATAAATAA 1782

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
              | | ||||| ||   || |||||||| || |  | |   |  | |   || || |
Db       1781 ATAAATTAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 1722

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGA 425
              ||||  ||  ||     |||      || | |  |     || |  | || | | | | |
Db       1721 ATAAATAAATAAATAAATAAATCGATATATAAATAAATAAATAAATAAATAAATAAATAA 1662

Qy        426 AAA 428
              | |
Db       1661 ATA 1659
```

RESULT 12
AAS61077/c
ID AAS61077 standard; DNA; 6109 BP.
XX
AC AAS61077;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #32.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 33; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour

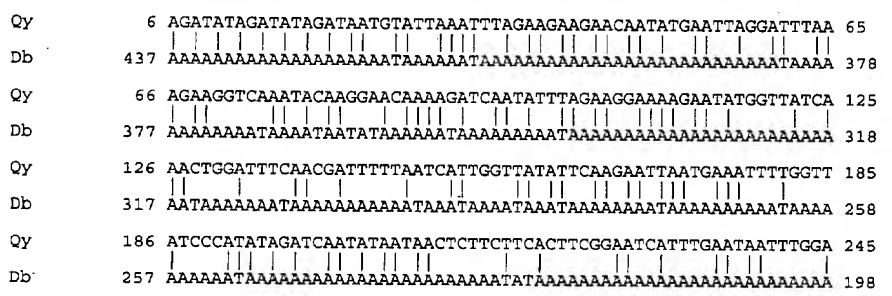
```
CC   dissimilar to cytosine, to enable analysis of cytosine methylations.
CC   The DNA sequences, oligomers (or sets/arrays) and method are
CC   useful in the diagnosis of diseases (or predisposition to diseases)
CC   associated with gene regulation and in therapy of such diseases, by
CC   enabling analysis of the cytosine methylation patterns of such genes,
CC   kits are provided. They are especially useful in diagnosis
CC   and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC   disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC   asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC   preeclampsia, graft versus-host disease. The present sequence is a
CC   sequence included in the sequence data for this specification and is
CC   associated with the human gene regulation-associated genes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6109 BP; 1232 A; 195 C; 1734 G; 2944 T; 4 other;

  Query Match             13.4%;  Score 59.8;  DB 24;  Length 6109;
  Best Local Similarity   46.3%;  Pred. No. 0.043;
  Matches  196;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       2081 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 2022

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       2021 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAACGAAAAAATAAATAAATAAAT 1962

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       1961 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAAAATAAATAAATAAAT 1902

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       1901 AAATAAATAAATAAATAAATATATAAATAAATAAATAATAAATAAATATATAAATAAATA 1842

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db       1841 AATAAATAAATAAATAAATATATAAATAAATAAATAAAAAATAAAAAAATAAATAAATAA 1782

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db       1781 ATAAATTAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 1722

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGA 425
Db       1721 ATAAATAAATAAATAAATAAATAAATCGATATATAAATAAATAAATAAATAAATAAATAA 1662

Qy        426 AAA 428
Db       1661 ATA 1659


RESULT 13
AAT41852
ID   AAT41852 standard; DNA; 9789 BP.
XX
AC   AAT41852;
XX
DT   20-FEB-1997  (first entry)
XX
DE   cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
XX
KW   Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW   detection; identification; treatment; prevention; parasite; ss.
XX
OS   Plasmodium falciparum MC type.
XX
FH   Key             Location/Qualifiers
FT   CDS             326..9497
FT                   /*tag= a
FT                   /product= Erythrocyte membrane protein
FT   misc_feature    518..520
FT                   /*tag= b
FT                   /transl_except= GTA encodes Tyrosine
FT   misc_feature    656..658
FT                   /*tag= c
FT                   /transl_except= ATT encodes Leucine
FT   misc_feature    2909..2911
FT                   /*tag= d
FT                   /transl_except= AAC encodes Aspartic acid
FT   misc_feature    3461..3463
FT                   /*tag= e
FT                   /transl_except= GAA encodes Glutamine
FT   misc_feature    5546..5548
FT                   /*tag= f
FT                   /transl_except= CCT encodes Arginine
FT   misc_feature    6254..6256
FT                   /*tag= g
FT                   /transl_except= AAT encodes Lysine
FT   misc_feature    6257..6259
FT                   /*tag= h
FT                   /transl_except= ATA encodes Tyrosine
FT   misc_feature    6263..6265
FT                   /*tag= i
FT                   /transl_except= AAC encodes Lysine
FT   misc_feature    6269..6271
FT                   /*tag= j
FT                   /transl_except= TTC encodes Isoleucine
FT   misc_feature    6272..6274
FT                   /*tag= k
FT                   /transl_except= ATA encodes Histidine
FT   misc_feature    6275..6277
FT                   /*tag= l
FT                   /transl_except= ATT encodes Asparagine
FT   misc_feature    6278..6280
FT                   /*tag= m
FT                   /transl_except= GGA encodes Tryptophan
FT   intron          7754..8478
FT                   /*tag= n
XX
PN   WO9633736-A1.
XX
PD   31-OCT-1996.
XX
PF   26-APR-1996;   96WO-US05798.
XX
PR   27-APR-1995;   95US-0430908.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Baruch DI,  Howard RJ,  Pasloske BL;
XX
DR   WPI; 1996-497376/49.
DR   P-PSDB; AAW00384.
XX
PT   New Plasmodium falciparum erythrocyte membrane proteins - used to
PT   develop products for the diagnosis, treatment or prevention of
PT   malaria parasite infections
XX
PS   Disclosure; Figure 12; 149pp; English.
XX
CC   A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC   membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection.  The polypeptides can inhibit, block or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC   probes and primers to identify a Plasmodium falciparum parasite, the
CC   primers used to generate characteristic amplification patterns from
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes the
```

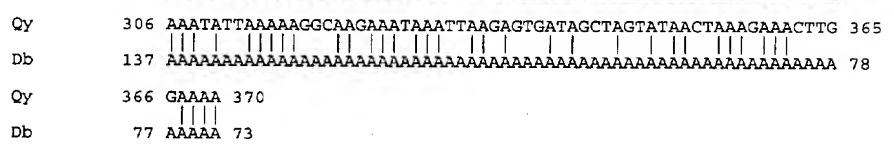
```
CC   PfEMP1 protein of the MC type of Plasmodium falciparum.  An
CC   alternative, truncated version of the coding sequence (a cDNA clone)
CC   is given in AAT41853.
XX
SQ   Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T; 0 other;

  Query Match             13.4%;  Score 59.8;  DB 17;  Length 9789;
  Best Local Similarity   47.1%;  Pred. No. 0.043;
  Matches  184;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       7888 AAAGAAATATAAAAAAAATTTATTAAAATAAATAAAAAAAAAAAAAATTCATAAAAAAAAA 7947

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       7948 AGAAATGAAAAATTTTATAAAAAAAAAATTATATTAATAAAAATAATGATTATAAAAAAA 8007

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       8008 ATTTATTAGAAATAAAATAAAAAAAAATTTATTAAATAAAAAAAAAAAGAAAAAAAAATG 8067

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       8068 TTAAAAAATATATATATATATATCATAAAATAAAAAAAAATGTTAAAAGAAATATATATA 8127

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db       8128 TTATAAAATAAAAAAAAAAATTTAAAAAAATGTTAAAAAAAAATATATATATATATATAA 8187

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db       8188 AATAAAAAAAAATAAAAAAAATGTAAAAAAAAAAAAAATTAAAAAAATAAAAAAATTAAA 8247

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAA 396
Db       8248 AAAAATAAAATAAAAAAAAAAAAATAAAAAAA 8278

RESULT 14
ABL34358/c
ID   ABL34358 standard; DNA; 12237 BP.
XX
AC   ABL34358;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2331.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
```

```
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2331; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 12237 BP; 3050 A; 207 C; 3024 G; 5956 T; 0 other;

  Query Match             13.4%;  Score 59.8;  DB 24;  Length 12237;
  Best Local Similarity   46.3%;  Pred. No. 0.043;
  Matches  196;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
Db       6985 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAAAAAAAAAA 6926

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
Db       6925 AAAAAAATAAATAAATCAACAAATAAATAAATAAATAAATAAATAAATAAAAAAAAAAAA 6866

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
Db       6865 AAAAAAATAAATAAATCAATAAATAAATAAATAAAATAAATAAATAAATAAATAAATAAA 6806

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
Db       6805 AAACAAAAAAAAAAATAAATAAATCAATAAATAAATAAATAAAATAAATAAATAAATAAA 6746

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
Db       6745 TAAATAAATAAATAAATAAAAAAAAAAAAAAAAAAAATAAATAAATCAACAAATAAATAAA 6686

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
Db       6685 TAAATAAATAAATAAATAAAAAAAAATAAATAAATAAATCAAAAATAAATAAATAAAAAAA 6626

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGA 425
Db       6625 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6566

Qy        426 AAA 428
Db       6565 ATA 6563

RESULT 15
ABX38235/c
ID   ABX38235 standard; cDNA; 446 BP.
XX
AC   ABX38235;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #3400.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
```

XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
PR 11-JAN-2000; 2000US-0480902.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX
PS Claim 2; SEQ ID No 3400; 245pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ Sequence 446 BP; 23 A; 8 C; 7 G; 408 T; 0 other;

Query Match 13.3%; Score 59.4; DB 25; Length 446;
Best Local Similarity 47.7%; Pred. No. 0.05;
Matches 174; Conservative 0; Mismatches 191; Indels 0; Gaps 0;





Search completed: February 16, 2004, 11:20:19
Job time : 281 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 10:54:01 ; Search time 2562 Seconds
(without alignments)
4240.477 Million cell updates/sec

Title: US-09-980-054A-11
Perfect score: 447
Sequence: 1 atgtcagatatagatataga.....accccgatgatatgtggtga 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	208.2	46.6	1059	29	CNS07EAO	AL441526 T3 end of
	2	115.6	25.9	902	29	CNS06VIL	AL417187 T3 end of
c	3	80.8	18.1	994	13	BX414650	BX414650 BX414650
c	4	78.2	17.5	1101	29	CNS001FB	AL060732 Drosophil
	5	76.4	17.1	1201	9	AL565455	AL565455 AL565455
c	6	75.6	16.9	1200	13	BX415878	BX415878 BX415878
	7	74.4	16.6	1101	29	CNS00EVL	AL069706 Drosophil
c	8	74	16.6	1201	13	BX462207	BX462207 BX462207
c	9	72.4	16.2	725	28	BH180166	BH180166 016_I_02-
	10	72.4	16.2	1201	13	BX439779	BX439779 BX439779
c	11	71.2	15.9	1201	13	BX335216	BX335216 BX335216
c	12	71	15.9	1201	29	CNS0167M	AL106396 Drosophil
	13	70.6	15.8	1101	29	CNS00EA9	AL068837 Drosophil
	14	70.2	15.7	997	14	CD049644	CD049644 AGENCOURT
	15	70	15.7	1001	29	CNS0155H	AL105023 Drosophil
c	16	69.8	15.6	942	29	CNS018GS	AL109318 Drosophil
	17	69.8	15.6	1201	9	AL543368	AL543368 AL543368
c	18	69	15.4	639	29	CNS038CX	AL232458 Tetraodon
	19	69	15.4	1126	13	BX446388	BX446388 BX446388
c	20	69	15.4	1164	29	CC218891	CC218891 CH261-14M
c	21	68.6	15.3	1044	13	BX415231	BX415231 BX415231
c	22	68.6	15.3	1139	28	AQ897537	AQ897537 HS_3153_A
c	23	68.6	15.3	1201	9	AL532464	AL532464 AL532464
	24	68	15.2	457	12	BJ333457	BJ333457 BJ333457
	25	68	15.2	885	29	AG136199	AG136199 Pan trogl
	26	67.8	15.2	455	12	BJ335466	BJ335466 BJ335466
	27	67.8	15.2	486	12	BJ394994	BJ394994 BJ394994
	28	67.6	15.1	1056	13	BX415058	BX415058 BX415058
	29	67.4	15.1	1201	13	BX355654	BX355654 BX355654
	30	67	15.0	945	29	CNS04D0K	AL285149 Tetraodon
c	31	67	15.0	1101	29	CNS016LI	AL106896 Drosophil
c	32	66.8	14.9	388	29	CNS039KI	AL234027 Tetraodon
	33	66.8	14.9	1201	9	AL547143	AL547143 AL547143
	34	66.6	14.9	865	28	BH180441	BH180441 016_H_20-
	35	66.6	14.9	865	29	CNS07MAN	AL617393 T3 end of
c	36	66.6	14.9	928	29	CNS00DKY	AL071865 Drosophil
	37	66.6	14.9	953	13	BX331505	BX331505 BX331505
c	38	66.6	14.9	1169	29	CNS06KHQ	AL402900 T3 end of
	39	66.6	14.9	1201	13	BX421282	BX421282 BX421282
c	40	66.4	14.9	903	29	CNS02WVM	AL217579 Tetraodon
	41	66.4	14.9	916	28	BH130447	BH130447 ENTOP15TF
	42	66.4	14.9	1101	29	CNS0021J	AL061936 Drosophil
c	43	66.4	14.9	1181	13	BX387370	BX387370 BX387370
	44	66.4	14.9	1201	9	AL532464	AL532464 AL532464
c	45	66.2	14.8	768	29	CNS01V5E	AL168683 Tetraodon

ALIGNMENTS

RESULT 1
CNS07EAO/c
LOCUS CNS07EAO 1059 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone XBD0AA002E07 of library XBD0AA from strain CBS 94 of Candida tropicalis, genomic survey sequence.
ACCESSION AL441526
VERSION AL441526.1 GI:12224752
KEYWORDS GSS.
SOURCE Candida tropicalis
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1059)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1059)
AUTHORS Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.

```
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 1059)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1059
                     /organism="Candida tropicalis"
                     /mol_type="genomic DNA"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="XBD0AA002E07"
                     /clone_lib="XBD0AA"
                     /note="end : T3"
     misc_feature    complement(<26. .>346)
                     /note="similar to Saccharomyces cerevisiae ORF YNL260c [
                     hypothetical protein ]"
                     /evidence=not_experimental
BASE COUNT      362 a    149 c    198 g    348 t      2 others
ORIGIN

  Query Match             46.6%;  Score 208.2;  DB 29;  Length 1059;
  Best Local Similarity   72.8%;  Pred. No. 1.3e-16;
  Matches  267;  Conservative    1;  Mismatches   99;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              |||   ||| | :| ||| || | ||||||||  ||||||||||||||||| | ||||||
Db        367 ATGAGCGATTTTKACATATATGAAGTATTAAACCTAGAAGAAGAACAATATAATTTAGGA 308

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              ||||| || || ||   |||    || ||||| ||||| ||||||||||| |||||||||
Db        307 TTTAAGGAGGGACAGGAACACTCTACCAAAGAACAATACTTAGAAGGAAAACAATATGGT 248

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              |||||||||||| |||||| | |||||||| |||| |||||||| | ||| |
Db        247 TATCAAACTGGGTTTCAAAGGTTTTTAATTGTTGGCTATATTCAGGGATTGGTAGGGGAA 188

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              |||||      || |||| | |||||||   ||     ||||||||     |||| | ||| |
Db        187 TGGTTGGATAATCTAGACAACTATAATGCTTCCAAATCACTTCAAGGTCATATAAATCAA 128

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              |    ||  | ||  |  | ||  |  | || ||||||||| |||||||||| | ||
Db        127 CTAAGTGAACTAATAACTGACATACCATTGACAAATGGAGATGAAGAAGTTGAAAAATAC 68

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              |||||||| || |||||||||||||||||| |||||||||| || |  ||||||||||||
Db         67 GAAAAAAACATCAAAAAGGCAAGAAATAAACTAAGAGTGATTGCAAACATAACTAAAGAA 8

Qy        361 ACTTGGA 367
              | |||||
Db          7 AATTGGA 1


RESULT 2
CNS06VIL
LOCUS       CNS06VIL                902 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T3 end of clone AX0AA021C12 of library AX0AA from strain CBS 7064
            of Pichia farinosa, genomic survey sequence.
ACCESSION   AL417187
VERSION     AL417187.1  GI:12198580
KEYWORDS    GSS.
SOURCE      Pichia farinosa
  ORGANISM  Pichia farinosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 902)
  AUTHORS   de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
  JOURNAL   FEBS Lett. 487 (1), 87-90 (2000)
  MEDLINE   20584725
   PUBMED   11152890
REFERENCE   3  (bases 1 to 902)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Pichia farinosa"
                     /mol_type="genomic DNA"
                     /strain="CBS 7064"
                     /db_xref="taxon:4920"
                     /clone="AX0AA021C12"
                     /clone_lib="AX0AA"
                     /note="end : T3"
BASE COUNT      273 a    162 c    166 g    296 t      5 others
ORIGIN

  Query Match             25.9%;  Score 115.6;  DB 29;  Length 902;
  Best Local Similarity   60.8%;  Pred. No. 2.4e-05;
  Matches  209;  Conservative    1;  Mismatches  125;  Indels    9;  Gaps    1;

Qy         13 GATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGT 72
              |||||||| ||||| ||  | ||||| || ||  | || ||    |||||  || |||||
Db        554 GATATAGAGAATGTGTTGGACTTAGAGGATGAGTACTACGATGAAGGATACAAGGAAGGG 613

Qy         73 CAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGA 132
              |||  |           | || |   |||| |||||||  |||||||||||||||||| ||
Db        614 CAAGGAGTCTTTGTTCAGGAGCTGCATTTGGAAGGAATGGAATATGGTTATCAAACAGGC 673

Qy        133 TTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG--------- 183
              | ||| ||||||||||  ||||||| ||  |||  |||||  | |  |||
Db        674 ATGCAAAGATTTTTAATAGTTGGTTACATGAAAGGGTTAATTGATTACTGGAAGAGCCAC 733
```

```
Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
Db        734 TTGTCTCAGTATGAGCAAATAGTTAGCATTAAGACACTAGAAAATCACTTAAGTCTGGCA 793

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303
Db        794 GAAGGCATTTTGGATGGCATTTCTATGGAWAATACGGAAGATGCTGTAAGAGAGTATGAA 853

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAG 347
Db        854 AAAGCCATACTAAAGAGCAAAAATAAAATTWGAGTAATAGCAAG 897
```

RESULT 3
BX414650/c

LOCUS	BX414650 994 bp mRNA linear EST 15-MAY-2003
DEFINITION	BX414650 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YN02 3-PRIME, mRNA sequence.
ACCESSION	BX414650
VERSION	BX414650.1 GI:30763455
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 994)
AUTHORS	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6015.f Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP001DG01NP1.

FEATURES Location/Qualifiers

```
     source          1. .994
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YN02"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      185 a    121 c     25 g    324 t    339 others
ORIGIN
```

Query Match 18.1%; Score 80.8; DB 13; Length 994;
Best Local Similarity 23.1%; Pred. No. 0.39;
Matches 77; Conservative 149; Mismatches 108; Indels 0; Gaps 0;

```
Qy         69 AGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAAC 128
Db        991 WKKWDDDDGRKKDWAWWAADKWAAARDRWWAGAARARRADWADDWAAAWRWWDRWTAWWA 932

Qy        129 TGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATC 188
Db        931 AKRKKWADWTTTKTTKKKDADAADWGDDWDWDATAWWWAADDWAWAARWAAATWWTDWDT 872

Qy        189 CCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGA 248
Db        871 AWWTRKDKATWKADDADDDAKTKTTTTTTTWWKAWTDWAGWTAKWWADATWADAADAWAA 812

Qy        249 TATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAA 308
Db        811 RAWWAARAWWRAAAAWWWAADADWWWWWWDWAADAAWWAAWDTKDKRAWRAADWARAAAA 752
```

```
Qy        309 TATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAA 368
Db        751 AAWWDWGRGGRADAAGKDKARAAWWDGAGGRDKDRDWAARWDDADAAARRAAAAAAATAW 692

Qy        369 AATTGATTCATTGGATAATTTGGTGAAAGAAGTA 402
Db        691 WATWWKWKDKDWAAAAAWDTAKGKGRKDDARDDW 658
```

RESULT 4
CNS001FB/c

LOCUS	CNS001FB 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL060732
VERSION	AL060732.1 GI:4939397
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

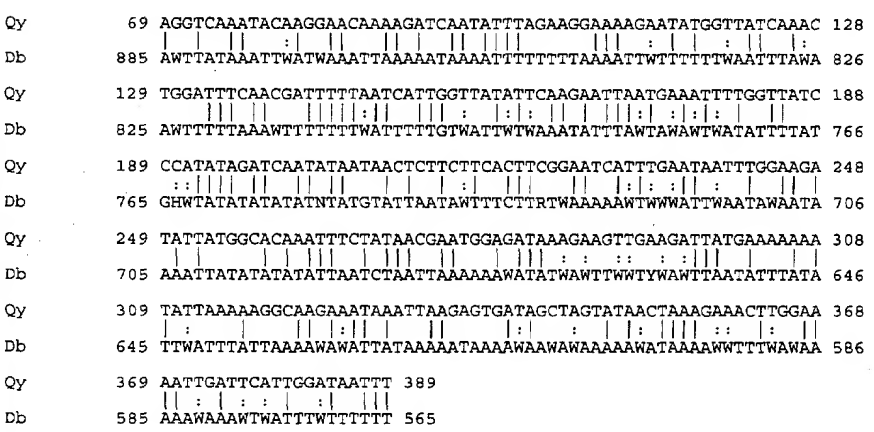
FEATURES Location/Qualifiers

```
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR04A23"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      288 a    110 c    103 g    491 t    109 others
ORIGIN
```

Query Match 17.5%; Score 78.2; DB 29; Length 1101;
Best Local Similarity 41.6%; Pred. No. 0.79;
Matches 168; Conservative 48; Mismatches 185; Indels 3; Gaps 1;

```
Qy          4 TCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTT 63
Db       1040 TAAAAAAAATATAAATWATTAAAAATAAAATWTAAAAAWATWAAAWAAAAAATTWAATAA 981

Qy         64 AAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTAT 123
Db        980 TTAWTATATTNAAAAAAAAWAAAAAAAAATTAATAAWWAAAAAAAAAAWATAATAATATA 921

Qy        124 CAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGG 183
Db        920 TAAAWWTAATWTAATWNATTWAAAAAATAA---TAWAATAWWAAATTAAWAAAAATAWAT 864

Qy        184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTG 243
```

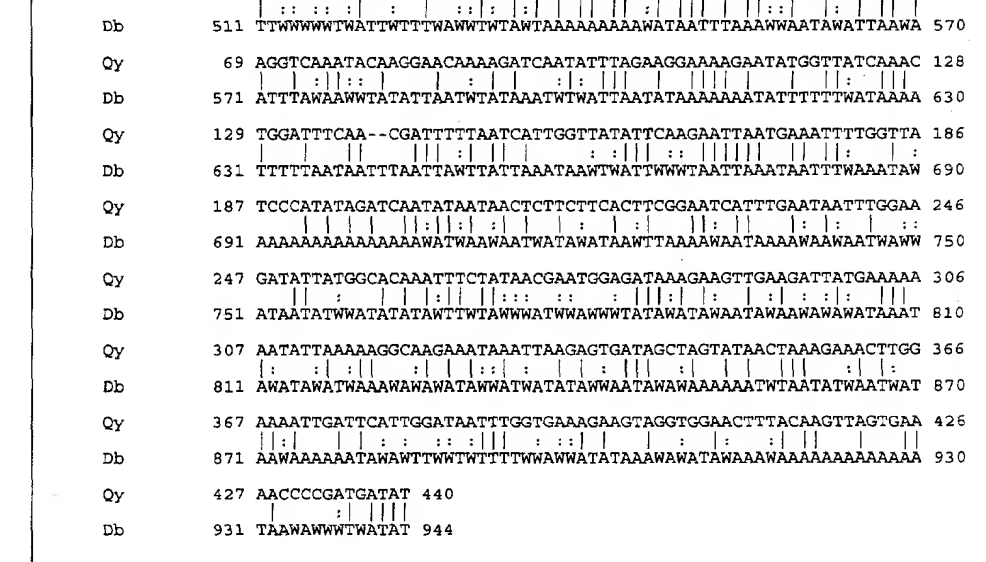
```
Db        863 WAAAAAWNTAWWTATAAWTTWAAAAATWTNTWAWWWAAAAATWWTTTTTTAATATATATW 804

Qy        244 GAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA 303

Db        803 TAATWTAATWTWWAAAAAAAATAWAAWAWTAWWTWAWWWNAWAAAWAAATAATATAWNAA 744

Qy        304 AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACT 363

Db        743 WAAAAATTWAWWANATWWWTWWWTAAAATAAAAAAAAAATAWATAAATAAWAAAAAAAAA 684

Qy        364 TGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGG 407

Db        683 AAAAAAAAAAWAWWAAAAAAAAAAGNNGGGGGGGGGGGGGGGG 640


RESULT 5
AL565455
LOCUS       AL565455               1201 bp    mRNA    linear   EST 12-MAY-2003
DEFINITION  AL565455 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF005YO18 3-PRIME, mRNA sequence.
ACCESSION   AL565455
VERSION     AL565455.2  GI:30549492
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12916848.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9232.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF005BH09NP1&cluster=9232.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF005BH09NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF005YO18"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      550 a     82 c     71 g    313 t    185 others
ORIGIN

  Query Match             17.1%;  Score 76.4;  DB 9;  Length 1201;
  Best Local Similarity   36.2%;  Pred. No. 1.3;
  Matches  155;  Conservative   71;  Mismatches  202;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60

Db        421 ATATCCTATATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 480

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120

Db        481 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 540

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180

Db        541 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 600

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240

Db        601 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 660

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300

Db        661 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 720

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360

Db        721 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 780

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420

Db        781 TATAWATAWATAWATAWATAWATAWATAWATATAWATAWATAWATAWATAWATAWATAWA 840

Qy        421 AGTGAAAA 428

Db        841 TATAWATA 848


RESULT 6
BX415878/c
LOCUS       BX415878               1200 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YI04
            5-PRIME, mRNA sequence.
ACCESSION   BX415878
VERSION     BX415878.1  GI:30765550
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP008BE02QP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP008YI04"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      428 a     70 c     85 g    447 t    170 others
ORIGIN

  Query Match             16.9%;  Score 75.6;  DB 13;  Length 1200;
  Best Local Similarity   41.2%;  Pred. No. 1.6;
  Matches  157;  Conservative   44;  Mismatches  180;  Indels    0;  Gaps    0;

Qy          9 TATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGA 68

Db        945 TTTATTWTWWAAAAAAWATTWAAATTWAAAATAAATAAAWAAAAAAWTTTTTSTAATTTT 886
```



RESULT 7
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 16.6%; Score 74.4; DB 29; Length 1101;

Best Local Similarity 36.6%; Pred. No. 2.3;
Matches 159; Conservative 76; Mismatches 197; Indels 2; Gaps 1;



RESULT 8
BX462207/c
LOCUS BX462207 1201 bp mRNA linear EST 22-MAY-2003
DEFINITION BX462207 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA clone CS0DG004YB03 5-PRIME, mRNA sequence.
ACCESSION BX462207
VERSION BX462207.1 GI:31023422
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 24.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DG004CA02QP1&cluster=24.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DG004CA02QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG004YB03"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed

```
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      308 a    141 c    153 g    522 t     77 others
ORIGIN

  Query Match           16.6%;  Score 74;  DB 13;  Length 1201;
  Best Local Similarity   44.9%;  Pred. No. 2.5;
  Matches  149;  Conservative   29;  Mismatches  154;  Indels    0;  Gaps    0;

Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              || : | | | | | ||| ||||    |::|||:| : || || || || |  ||  | |
Db        900 ATTYTAAAAAAAAAAATAAATAAATAAWWAAAWTAWAAAAAAAAAAAAAAAAAAAAAAA 841

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
                ||| ||   ||| | ||  || |||| | :||:|   | ||  |||| | | |
Db        840 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAYAAWAAAAAAAAAAAAAAAAATTTTTAA 781

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              |  |   |  |    ||  |     ||:  ||    | |   |: ||  |:| ||:|||
Db        780 AAAAATTTTWWAAAAWAAAAAAAAAAAAWWTTTTTAAAAAAAAAWAAAAAAWTAAAWTTT 721

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              |    |:  | |:  :: || | || |:   ||||   ||   :| |:   |: :::
Db        720 TTAAAAWAWAAAAWTWWWAAAAAAAAAAWTAWTTTTTTTTTTTTTWWAAAAWAAAAWAWWW 661

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              ||   | | ||| |   : |||       |: :  |:  | | ||  ||    || | |
Db        660 TTWWTAAAAATTTTTWTWAAAAAAAAAAWTWAAAAWAAAAAAAAATWAAAAAAAAAAAAAA 601

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATT 332
              ||||||| | ||||   || ||| ||:::
Db        600 AAAAAAAAAAAAAAAAAAWWAAAAAAAAWWW 569


RESULT 9
BH180166/c
LOCUS       BH180166                 725 bp    DNA     linear   GSS 19-OCT-2001
DEFINITION  016_I_02-21 SmBAC1 Schistosoma mansoni genomic clone 016I02 5',
            genomic survey sequence.
ACCESSION   BH180166
VERSION     BH180166.1  GI:16281799
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams
            ,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
COMMENT     Other_GSSs: 016_I_02-rev
            Contact: Pierce RJ
            INSERM U 167
            Institut Pasteur de Lille
            1 rue du Professeur A. Calmette, 59019-Lille, France
            Tel: (33) (0)3 20877783
            Fax: (33) (0)3 20877888
            Email: Raymond.Pierce@pasteur-lille.fr
            CNS sequencing ID=DG0AA016BE01CP1
            Plate: 016  row: I  column: 02
            Seq primer: M13 -21 primer
            Class: BAC ends
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Schistosoma mansoni"
                     /mol_type="genomic DNA"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="016I02"
                     /sex="mixed"
                     /dev_stage="cercariae"
                     /lab_host="Biomphalaria glabrata"
                     /clone_lib="SmBAC1"
                     /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                     Hind III digested and size-selected S. mansoni  cercarial
                     DNA was ligated into Hind III digested pBeloBAC 11 vector
                     and used to transform E. coli DH10B. The complete library
                     contains 23808 clones from 4 independent
                     sizing-ligation-transformations.  Average insert size
                     ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT      151 a     52 c     39 g     420 t     63 others
ORIGIN

  Query Match           16.2%;  Score 72.4;  DB 28;  Length 725;
  Best Local Similarity   42.8%;  Pred. No. 4.6;
  Matches  184;  Conservative   30;  Mismatches  216;  Indels    0;  Gaps    0;

Qy         14 ATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAAGAAGGTC 73
              | | | ||  | |:|   |   | ||| |: || || |  ||  |  |  ||| ||
Db        564 AAASAMATTCTTTMTMSGAAAAAWAAWAWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 505

Qy         74 AAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAAACTGGAT 133
              :|| | ||  |  :||| |  || |  ||      |||| || || | |  ||:    |:
Db        504 WAAAAAAAAAATAWAAAAAAAAAAAAAAATAWTTAAAAAAAAAAATWATAAAAAAWAAAAAW 445

Qy        134 TTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTATCCCATA 193
              : || : :  :|| |   |  | |  || ||: ||  :|||      |    : |
Db        444 AWAAAAWWAWAAWAARAAAAAWAAAAAAAAAAAAAAWAAAAAWAATRAAAAAAAAAAAWAA 385

Qy        194 TAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAAGATATTA 253
              |  || || : || :  : |      |      ||  ||   || || |  :|| ||:
Db        384 TTWATAAAGWGAAARTAWATAWAAAAAAAAAAAAAAAAATAAAAAAAAATAARAAAATWAAG 325

Qy        254 TGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAAAATATTA 313
              |     || :  : ||   ||      ||:| : | :| ||  | |  || ||||:|| |
Db        324 TWWTTRGAAAWAAWTTATTTAAGTATAATRATRTATWTAAAATTAAATAATAAAAWATAA 265

Qy        314 AAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGGAAAATTG 373
              ||||   || ||| ||   ||  ||||||   | ||| :  ||: ||| |  | :|| |:
Db        264 AAAAAAAAAAAAAGAAGAAAAATGTGATAATAAATATTMGGAARTAAAGAAGAARAAGTR 205

Qy        374 ATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTAGTGAAAACCCCG 433
              ||  | | ||      || ||   || || | | |  |  || | |:||
Db        204 ATGAAGGGAATGGAAGAGTAAATATAGAAGTTTGTAATGATRATATTGTTAARAARAAAA 145

Qy        434 ATGATATGTG 443
              ||:| |  ||
Db        144 ATRAAARATG 135


RESULT 10
BX439779
LOCUS       BX439779                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX439779 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014YF05
            3-PRIME, mRNA sequence.
ACCESSION   BX439779
VERSION     BX439779.1  GI:30771778
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
```

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3370.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE014CC03NP1&cluster=3370.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DE014CC03NP1.

FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE014YF05"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 455 a 150 c 90 g 299 t 207 others
ORIGIN

Query Match 16.2%; Score 72.4; DB 13; Length 1201;
Best Local Similarity 38.6%; Pred. No. 3.9;
Matches 165; Conservative 64; Mismatches 196; Indels 3; Gaps 1;

```
Qy          1 ATGTCAGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGA 60
              :|:|   : : | | :|   | ||   | ||||||: |: || :| | ||| ||     |
Db        669 WTKTTTTWAWAATAAWTTTTTTATTATTAAAATTTWTAWAAAAWAAATTATTAAAATWAA 728

Qy         61 TTTAAAGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGT 120
              :| |||      :|| | ||   | : || ||  | ||:|| |   :|||   ||   :
Db        729 WTAAAAATTWTAHAATAAAAATWTAAWTAAAATATAAATKTAWATTTWAAAATTTAATTW 788

Qy        121 TATCAAACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTT 180
              | | |||    |||| :|  |||| |||    :  ||  || | :|| |:| :|:|||
Db        789 TTTTAAATATAATTTAWAAAATTTATAAMATAWTTTT---TTWATAWATAAWTTWAWTTT 845

Qy        181 TGGTTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAAT 240
              |    :|   ::|:| |  |: |:   | |   :|  :|:|  |:|:
Db        846 TAAAAWTAAAWWAWAAAAAAWAAWTTTTAAAAWWWTWWAATAWTTWAATWTWTTTAWAWA 905

Qy        241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTAT 300
              :   :   ||::||   | |: |:   :|    ||:  | | ||| || ::  : |::
Db        906 WAATWTTTTAWWATAWTAAAWTTWAWAWTTTTAAAWTWATAAAAAWAATWWATWTAWWTA 965

Qy        301 GAAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
              |||||:||   |||:    ||  |: ||| || |   | |   | :|: :  |||
Db        966 TAAAAAWATTAAAAAWTWWAAATTAWAAAAAAAAAAAWAAAAAAAAATWAWTWTAAAAATT 1025

Qy        361 ACTTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTT 420
              :      : ||| |   :::   | :||   | | : || |:  :  || |: | ||
Db       1026 WAAAWATWTATTWAAAAWWWATTTTWTTTWTTTATWWAAWTWAAWAAAAWTWAAAAAAA 1085

Qy        421 AGTGAAAA 428
              |   :|||
Db       1086 AAAAWAAA 1093
```

RESULT 11
BX335216/c
LOCUS BX335216 1201 bp mRNA linear EST 02-MAY-2003
DEFINITION BX335216 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI012YG10 5-PRIME, mRNA sequence.
ACCESSION BX335216
VERSION BX335216.1 GI:30341402
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 714.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI012BD05QP1&cluster=714.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI012BD05QP1.

FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI012YG10"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 391 a 115 c 137 g 453 t 105 others
ORIGIN

Query Match 15.9%; Score 71.2; DB 13; Length 1201;
Best Local Similarity 43.6%; Pred. No. 5.4;
Matches 173; Conservative 33; Mismatches 188; Indels 3; Gaps 1;

```
Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              | | ||| | ||| ||| | |||||||: ||  | || |  || :  :|     |   ||
Db       1169 AAAAATAAAAATATATATTTTATTAAAWATATTATAATATAAAAWAAWAAATAAAAAAAA 1110

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
              | |   | ||:| |:   | : ||    | | : || ::   | | ||| |   :|| |
Db       1109 ANATWTTATAAWATAWAATAAWNAAATATATTTWATATWWTATATANAATTTATAWATAA 1050

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
              ||    |::| ||    | | |::   :  |:    : || |::|: | :|  | |  ||
Db       1049 AA---AAWWTAAATATATATAAWWATAWWATWTATAWAAATAWWTWTTAWATATATTTTT 993

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
              | | |  |||:|| ||  | |  | ||   ||  |  : |: ||:|  | |
Db        992 NAATATAAAATAAATATAWAAAAATATATATATATTTWAAAAAAAAAWWAWAAAWATATA 933

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305
              | :   ||    : | |||| | ||   ||  | |||:      | ||  ||   ||||
Db        932 AWWATATAAATWWTATATTTNAAAAATATATATATATAWTATTAATTAATTTTNATAAAA 873

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365
              |||:|  |||||   ||    |||   ||     |:|  ||  : :|  ||| |||
Db        872 AAAWAAAAAAAAWANAAATTTTAATAAAAATAAWAWAAATATAWAWAAAAAAAAAAAAAA 813

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTA 402
              ||||   ||| |||  |:||  |  | ||| || ||
Db        812 AAAAAAAAATTAATTATAWAAAATTATAAAAAAATTA 776
```

RESULT 12
CNS0167M/c
LOCUS CNS0167M 1201 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106396
VERSION AL106396.1 GI:5621701

```
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15M24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      323 a     87 c     79 g    551 t    161 others
ORIGIN

  Query Match             15.9%;  Score 71;  DB 29;  Length 1201;
  Best Local Similarity   41.8%;  Pred. No. 5.7;
  Matches  175;  Conservative   42;  Mismatches  199;  Indels    3;  Gaps    1;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65
              |||||| | ||| |||:   ||  |  : | ||      | || |: || ||  || |||
Db       1180 AAATATAAAAATAAATAWAWAATATATAWNAAAAANTATAAAAAAWAAAATAWWATATAA 1121

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125
              | ||    ::::| |    || |:|| || | :|   | ||  |||| |:::|  :||  |
Db       1120 AWAAWAAAWWWWATATTAAATAWAATATAANWAAAAAAAAAAAAAAAAAWWWTTTHTANAA 1061

Qy         126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
               | |   ||:  |  :  |:|   |  ||  || | |  |: :|  :|  |||   | |
Db       1060 TATTTWTTTWNTATAWAATWTTTTTTTTTTTTTTTATAWAWAAAWAAAAAAAAATTTTA 1001

Qy         186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245
               |    | |:  || | :| |||   |    |   |:   :  |  |:  || ||     |
Db       1000 AAAATAAAWTAATTATWAAAATTTTTAAAAATTTTTWATWWTTTTTWAAAAAAAAAAWAA 941

Qy         246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAA-- 303
                 :|  :|        ||:| || |:  |:|     ::   ||     || :|||  ||
Db       940 TATWAAAWTTTTTTTTATWTATAAAWAWAWTTTTTTYWTTAAAAAAAAAAWTTAAAAATW 881

Qy         304 -AAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAAC 362
               |||||| |||||    |: ||| |||  || |   ::|   | | |:| || | :|
Db       880 TAAAAATTATAAAATAAAAWAAAAAAAAAAAAAAAAAAAWWAAAAAWTTTWATTATAAWATT 821

Qy         363 TTGGAAAATTGATTCATTGGATAATTTGGTGAAAGAAGTAGGTGGAACTTTACAAGTTA 421
               ||  |||| |  |||  ||  |||    ||| || |    :| :: | || |::
Db       820 TTTWAAAAAAAAAAAATTTAATTYTTTTTNAAAATAAAAAAWAAAWAWWWAAAAAATWW 762


RESULT 13
CNS00EA9
LOCUS       CNS00EA9               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR28K21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL068837
VERSION     AL068837.1  GI:4949081

KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR28K21"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      502 a      8 c    151 g    297 t    143 others
ORIGIN

  Query Match             15.8%;  Score 70.6;  DB 29;  Length 1101;
  Best Local Similarity   45.0%;  Pred. No. 6.6;
  Matches  175;  Conservative   29;  Mismatches  183;  Indels    2;  Gaps    1;

Qy          7 GATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAAA 66
              || | | ||:|   | | | |   ||| |  || ||:||   ||| |  :|  || ||| |
Db        264 GAAAAAAATWTGTGTTAAGGAAAGAATATTAAAAAARAATTTTATAATAWAWTATGTATA 323

Qy         67 GAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCAA 126
              ||| | |||:| ||   | :| | :  || |  :|  | |||||:|| || ||    ||
Db        324 GAATATWAAAWAAAAAWTAAWATAAWAAAARAAAWAAGAWGAAAARAAGATTGTATAGAA 383

Qy         127 ACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTTA 186
               |      ||   | |  |:  | :| ||||    |: :||::|  |||        :
Db        384 AAGAATATTGTTTGGGTAAAAWAAAAWGTTTATGAGTAWAWATWWAAAAAAAAAAAWAAW 443

Qy         187 TCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGAA 246
               :   : ||   : | |  | |||    |    | ||   |:: | | || |  | |  |
Db        444 WATTWAATTATWAATTTAATTAAAATATWAATATTTAAAAWWAAATGGATTTTTGGGTTA 503

Qy         247 GATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAAA 306
               ||| |||:    :   |  | :  ||   |  ||  |||   |:|||||  :||||
Db        504 AATAATATRAWTAWGTATATGAATWTAAAAAAAAGAAAGAAAGGAAAWGATTAAAWAAAA 563

Qy         307 AATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTGG 366
               | ||||:|    || | ||   | | |:    |  ||  | || | || ||  ||
Db        564 TAAATTAMATTATTAAAATATGTTTAAGGRTAATAAAGTAAAAAAATTGAAAGAATATG- 622

Qy         367 AAAATTGATTCATTGGATAATTTGGTGAA 395
               |||||||| | || |||| : | |||:
Db        623 -AAATTGATTTAATGTATAAAWGGTTGAW 650


RESULT 14
CD049644
```

```
LOCUS       CD049644                 997 bp    mRNA    linear   EST 09-MAY-2003
DEFINITION  AGENCOURT_13986201 NIH_MGC_172 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CD049644
VERSION     CD049644.1  GI:30485777
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jamie Thompson, University of WI
             cDNA Library Preparation: Gina Zastrow-Hayes
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDKM47  row: e  column: 22
            High quality sequence start: 10
            High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source          1..997
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cell"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_172"
                     /note="Vector: pDONR201; Site_1: attP2; Site_2: attP1;
                     LIBR_PRIMING - oligo dT; METHOD - full-length enriched;
                     Embryonic Stem Cells H1; LIBR PROVIDER - Bradfield"
BASE COUNT      427 a    128 c    113 g    306 t     23 others
ORIGIN

  Query Match             15.7%;  Score 70.2;  DB 14;  Length 997;
  Best Local Similarity   46.8%;  Pred. No. 7.6;
  Matches  183;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65

Db        393 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 452

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125

Db        453 AAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 512

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185

Db        513 AAAANNNNNNNNNNNNNNNTTATTTAAAAAAAAATTTTATTTATAAAAAAAAATTTTTTT 572

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245

Db        573 TTATATTTAATTTNTAAAAAAAAATTTTTTTTTAATTTTTAATATTAAAATTTTTTATTT 632

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305

Db        633 TTTTTTNAATATATATATTTTTTTTATTAAATAATAAAATAAAATTTATATATTTTTATA 692

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365

Db        693 AATTAATAATNAAATAAAAAATTTATTTATTTTNTAAAATAAATAAAAAATTATTTTTTT 752

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAA 396

Db        753 TATTATATATTAATTATAAAATAATTTAAAA 783

RESULT 15
CNS0155H
LOCUS       CNS0155H                1001 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN13C23 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL105023
VERSION     AL105023.1  GI:5617037
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..1001
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN13C23"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      266 a    219 c    134 g    150 t    232 others
ORIGIN

  Query Match             15.7%;  Score 70;  DB 29;  Length 1001;
  Best Local Similarity   34.0%;  Pred. No. 8.1;
  Matches  134;  Conservative   80;  Mismatches  180;  Indels    0;  Gaps    0;

Qy          6 AGATATAGATATAGATAATGTATTAAATTTAGAAGAAGAACAATATGAATTAGGATTTAA 65

Db        606 AARKTKWGARGGAAAAAAAGWDADDAATWMRKAARAAKADADWAAAAACAAAANVHNAAW 665

Qy         66 AGAAGGTCAAATACAAGGAACAAAAGATCAATATTTAGAAGGAAAAGAATATGGTTATCA 125

Db        666 AAARGKAAARRRAAAAAAAAARAGDAARAAAAWGRAAARDARAAAAAATAAAWWAATAARW 725

Qy        126 AACTGGATTTCAACGATTTTTAATCATTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185

Db        726 AARARDARAAAAAAATTAWWTAAWWWAWAAWWWWAAATAAAAADTTWTATWAAATAAAAA 785

Qy        186 ATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGAATAATTTGGA 245

Db        786 AAWWTAAAAAAATATTTAWAWAWAAAAAATWAAAAATTWWAAAAWWTAWWAAAAAAAATTW 845

Qy        246 AGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAAGTTGAAGATTATGAAAA 305

Db        846 AATTTTTTAAAWAWAAWWAATWAWATATAAAADAAAAAAAAAAWAAWATAAAATTWAAAA 905

Qy        306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATAACTAAAGAAACTTG 365

Db        906 WTTWWWTAWAWADWTWTTTWTTWTWWAAWAAATAAWTWTAWTTAATWTAAWAATAAAWAT 965

Qy        366 GAAAATTGATTCATTGGATAATTTGGTGAAAGAA 399

Db        966 AAAAAATWAWAAWWAWATAAWWTAAAAAAAAAWA 999

Search completed: February 16, 2004, 12:51:31
```

Job time : 2569 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:27:58 ; Search time 2528 Seconds
(without alignments)
1422.889 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980054/runat_13022004_163840_26309/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054_@CGN_1_1_2810_@runat_13022004_163840_26309 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	453	58.7	1059	29	CNS07EAO	AL441526 T3 end of
2	296.5	38.4	902	29	CNS06VIL	AL417187 T3 end of
c 3	200.5	26.0	609	29	BZ296387	BZ296387 CG2026.f1
4	196.5	25.5	433	29	BZ300548	BZ300548 KD0783.R1
c 5	187	24.2	1030	29	CNS076JA	AL431468 T7 end of
6	138.5	17.9	849	29	CNS06PHM	AL409376 T7 end of
c 7	130	16.8	784	14	CB632715	CB632715 OSIIEb11C
8	122	15.8	606	13	BQ461520	BQ461520 HD04B05r
9	122	15.8	611	13	BU968430	BU968430 HB07H22r
10	122	15.8	641	13	BQ472468	BQ472468 HB09H22T
11	114	14.8	530	9	AI994649	AI994649 701499129
c 12	110.5	14.3	879	29	BZ992706	BZ992706 PUEBB09TD
13	110	14.2	477	14	CA764105	CA764105 AF53-Rpf_
14	110	14.2	566	9	AV913233	AV913233 AV913233
15	110	14.2	648	14	CB632714	CB632714 OSIIEb11C
c 16	109.5	14.2	941	29	CC405283	CC405283 PUHTB67TB
17	108	14.0	641	9	AJ432075	AJ432075 AJ432075
18	108	14.0	772	10	BF627938	BF627938 HVSMEb000
19	107	13.9	630	14	CD431595	CD431595 ETH1_9_C0
20	105.5	13.7	637	10	BF054068	BF054068 EST439298
21	105.5	13.7	694	14	CD038635	CD038635 UTPPI010_
22	105	13.6	409	28	AQ962576	AQ962576 LERGG28TR
c 23	105	13.6	571	14	CB885171	CB885171 3529_1_82
c 24	104	13.5	661	13	BQ115339	BQ115339 EST600915
25	102.5	13.3	533	10	BE237388	BE237388 146664 MA
26	102.5	13.3	632	10	BG095637	BG095637 EST460192
27	102.5	13.3	645	10	BG599966	BG599966 EST504861
28	102.5	13.3	650	12	BM405065	BM405065 EST579392
29	102.5	13.3	689	13	BQ115338	BQ115338 EST600914
30	102.5	13.3	720	12	BI432791	BI432791 EST535552
31	101.5	13.1	688	10	BG616931	BG616931 602615605
32	100.5	13.0	366	9	AA304792	AA304792 EST175780
33	100	13.0	569	14	W44346	W44346 zc27b08.r1
34	99.5	12.9	458	9	AA148771	AA148771 zl04f04.r
35	99.5	12.9	497	10	BF654530	BF654530 278745 MA
36	99.5	12.9	506	12	BM712319	BM712319 UI-E-DW1-
37	99.5	12.9	533	10	BF815233	BF815233 RC5-CI014
38	99.5	12.9	564	13	BU662888	BU662888 c192g02.z
39	99.5	12.9	565	12	BM761310	BM761310 K-EST0042
40	99.5	12.9	585	13	BX474554	BX474554 DKFZp686K
41	99.5	12.9	589	12	BM849788	BM849788 K-EST0130
42	99.5	12.9	607	10	AW959687	AW959687 EST371757
43	99.5	12.9	687	9	AV682155	AV682155 AV682155
44	99.5	12.9	712	12	BM767075	BM767075 K-EST0049
45	99.5	12.9	716	12	BM850088	BM850088 K-EST0130

ALIGNMENTS

RESULT 1
CNS07EAO/c
LOCUS CNS07EAO 1059 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone XBD0AA002E07 of library XBD0AA from strain CBS 94 of Candida tropicalis, genomic survey sequence.
ACCESSION AL441526
VERSION AL441526.1 GI:12224752
KEYWORDS GSS.
SOURCE Candida tropicalis
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1059)

```
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1059)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 1059)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1059
                     /organism="Candida tropicalis"
                     /mol_type="genomic DNA"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="XBD0AA002E07"
                     /clone_lib="XBD0AA"
                     /note="end : T3"
     misc_feature    complement(<26. .>346)
                     /note="similar to Saccharomyces cerevisiae ORF YNL260c [
                     hypothetical protein ]"
                     /evidence=not_experimental
BASE COUNT      362 a    149 c    198 g    348 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              7.12e-45        Length:         1059
Score:                  453.00          Matches:        85
Percent Similarity:     81.15%          Conservative:   14
Best Local Similarity:  69.67%          Mismatches:     23
Query Match:            58.68%          Indels:         0
DB:                     29              Gaps:           0

US-09-980-054A-12 (1-148) x CNS07EAO (1-1059)

Qy          1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGly 20
              ||||||||  |||   ||||||||||||||||||||||||||||  ||||||
Db        367 ATGAGCGATTTTKACATATATGAAGTATTAAACCTAGAAGAAGAACAATATAATTTAGGA 308

Qy         21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
              ||||||||||||||    ||||||:::||||||||||||||||:::||||||
Db        307 TTTAAGGAGGGACAGGAACACTCTACCAAAGAACAATACTTAGAAGGAAAACAATATGGT 248

Qy         41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
              ||||||||||||||||||||||||:::||||||||||||   |||:::
Db        247 TATCAAACTGGGTTTCAAAGGTTTTTAATTGTTGGCTATATTCAGGGATTGGTAGGGGAA 188

Qy         61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
              ||||||   ::::::|||   ||||||   |||   ||||||:::   |||:::|||
Db        187 TGGTTGGATAATCTAGACAACTATAATGCTTCCAAATCACTTCAAGGTCATATAAATCAA 128

Qy         81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
              |||   :::::::::      |||   :::||||||||||||:::||||||||||   |||
Db        127 CTAAGTGAACTAATAACTGACATACCATTGACAAATGGAGATGAAGAAGTTGAAAAATAC 68

Qy        101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db         67 GAAAAAAACATCAAAAAGGCAAGAAATAAACTAAGAGTGATTGCAAACATAACTAAAGAA 8

Qy        121 ThrTrp 122
                 |||
Db          7 AATTGG 2

RESULT 2
CNS06VIL
LOCUS       CNS06VIL                 902 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T3 end of clone AX0AA021C12 of library AX0AA from strain CBS 7064
            of Pichia farinosa, genomic survey sequence.
ACCESSION   AL417187
VERSION     AL417187.1  GI:12198580
KEYWORDS    GSS.
SOURCE      Pichia farinosa
  ORGANISM  Pichia farinosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 902)
  AUTHORS   de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
  JOURNAL   FEBS Lett. 487 (1), 87-90 (2000)
  MEDLINE   20584725
   PUBMED   11152890
REFERENCE   3  (bases 1 to 902)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Pichia farinosa"
                     /mol_type="genomic DNA"
                     /strain="CBS 7064"
                     /db_xref="taxon:4920"
                     /clone="AX0AA021C12"
                     /clone_lib="AX0AA"
                     /note="end : T3"
```

```
BASE COUNT      273 a    162 c    166 g    296 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:                8.28e-26       Length:          902
Score:                    296.50         Matches:         58
Percent Similarity:       71.55%         Conservative:    25
Best Local Similarity:    50.00%         Mismatches:      30
Query Match:              38.41%         Indels:          3
DB:                       29             Gaps:            1

US-09-980-054A-12 (1-148) x CNS06VIL (1-902)

Qy          5 AspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGly 24
              ||||||:::|||||||||||:::||||||:::|||   |||:::  |||:::|||||||||
Db        554 GATATAGAGAATGTGTTGGACTTAGAGGATGAGTACTACGATGAAGGATACAAGGAAGGG 613

Qy         25 GlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
              |||             ::::::  :::||||||||  ||||||||||||||||||||
Db        614 CAAGGAGTCTTTGTTCAGGAGCTGCATTTGGAAGGAATGGAATATGGTTATCAAACAGGC 673

Qy         45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHis 64
                |||||||||||||||:::||||||::::::  |||:::  :::|||  ||||||
Db        674 ATGCAAAGATTTTTAATAGTTGGTTACATGAAAGGGTTAATTGATTACTGGAAGAGCCAC 733

Qy         65 IleAspGlnTyrAsn---------AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeu 81
              :::  ||||||          :::      :::|||  |||||||||:::
Db        734 TTGTCTCAGTATGAGCAAATAGTTAGCATTAAGACACTAGAAAATCACTTAAGTCTGGCA 793

Qy         82 GluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGlu 101
              |||  |||:::      ||||||:::  |||  :::      |||  :::||||||
Db        794 GAAGGCATTTTGGATGGCATTTCTATGGAWAATACGGAAGATGCTGTAAGAGAGTATGAA 853

Qy        102 LysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
              |||  |||  |||::::::||||||:::  |||||||||||:::
Db        854 AAAGCCATACTAAAGAGCAAAAATAAAATTWGAGTAATAGCAAGCCTC 901

RESULT 3
BZ296387/c
LOCUS       BZ296387                609 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  CG2026.f11 Candida glabrata Random Genomic Library Candida glabrata
            genomic clone CG2026, genomic survey sequence.
ACCESSION   BZ296387
VERSION     BZ296387.1  GI:24438257
KEYWORDS    GSS.
SOURCE      Candida glabrata
  ORGANISM  Candida glabrata
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG2026"
                     /clone_lib="Candida glabrata Random Genomic Library"
BASE COUNT      183 a    127 c     85 g    214 t
ORIGIN

Alignment Scores:
Pred. No.:                2.78e-14       Length:          609
Score:                    200.50         Matches:         43
Percent Similarity:       57.64%         Conservative:    40
Best Local Similarity:    29.86%         Mismatches:      50
Query Match:              25.97%         Indels:          11
DB:                       29             Gaps:            3

US-09-980-054A-12 (1-148) x BZ296387 (1-609)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              ::::::   |||   :::|||   |||||||||:::   |||:::  |||:::::|||
Db        591 ATGAATACTGATGAATTACTATTCTTAGAGGAACAGTTCTATCAAGAAGGTTATGAAGAA 532

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              |||:::  :::      |||:::     :::|||||||||||:::|||   |||
Db        531 GGCAGAAATGAAAACTTGAAGAACAATTTAATAGAAGGTAAAGAGTTTGGGTTACAAGTT 472

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              |||||||||||||:::::::::|||  :::      :::          |||
Db        471 GGATTTCAAAGATTTGTGTTACTAGGTCAAATGGTTGGGATGTGTGATGTC---CTTGAT 415

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                 :::      ||||||     ||||||     ::::::  |||:::  :::
Db        414 TCTTTACAGTTGAAGAATAAT------TCGTTATTTAAGAACATCGCAAATGTCAGAGAG 361

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              :::::::::  |||  :::  |||  ::::::  ||||||     ||||||
Db        360 CTGATCAGCTCTATCGAGCTTAACAACAAAGAGGAGAATGTTGAAAGTTTAGAAAAGACT 301

Qy        104 IleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLys 123
              :::  |||  :::|||||  |||:::|||          :::      :::|||
Db        300 CTAGTGAAATTAAAGAATAAATTTAGATTGATTTTACTTTCATTCCAGAGGGAATTTAAA 241

Qy        124 ----------------------IleAspSerLeuAspAsnLeuValLysGluValGly 135
                                   |||:::  ::::::|||  :::  :::
Db        240 CACAACGATCGCAAGCCTATCAATTTTGATAATTTTGAAGACTTGTCAAGAGCAATTGCA 181

Qy        136 GlyThrLeuGln 139
              |||   ::::::
Db        180 GGAGAAATCAAA 169

RESULT 4
BZ300548
LOCUS       BZ300548                433 bp    DNA     linear   GSS 31-OCT-2002
DEFINITION  KD0783.R1 Kluyveromyces delphensis Random Genomic Library
            Kluyveromyces delphensis genomic clone KD0783, genomic survey
            sequence.
ACCESSION   BZ300548
VERSION     BZ300548.1  GI:24445449
KEYWORDS    GSS.
SOURCE      Kluyveromyces delphensis
  ORGANISM  Kluyveromyces delphensis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
  TITLE     Evidence from comparative genomics for a complete sexual cycle in
            the 'asexual' pathogenic yeast Candida glabrata
  JOURNAL   Genome Biol. 4 (2), R10 (2003)
  MEDLINE   22508158
   PUBMED   12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
```

```
     source          1. .433
                     /organism="Kluyveromyces delphensis"
                     /mol_type="genomic DNA"
                     /strain="CBS 2170"
                     /db_xref="taxon:51657"
                     /clone="KD0783"
                     /clone_lib="Kluyveromyces delphensis Random Genomic
                     Library"
BASE COUNT      161 a     54 c     87 g    131 t
ORIGIN

Alignment Scores:
Pred. No.:              5.47e-14        Length:         433
Score:                  196.50          Matches:        41
Percent Similarity:     60.33%          Conservative:   32
Best Local Similarity:  33.88%          Mismatches:     43
Query Match:            25.45%          Indels:         5
DB:                     29              Gaps:           2

US-09-980-054A-12 (1-148) x BZ300548 (1-433)

Qy          4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
              :::|||:::|||   :::|||   |||||||||:::   |||:::   |||:::::|||
Db         49 ATGGATGTTGATGAGTTGTTATTTTTAGAGGAGAAGTTTTACAAAGAAGGCTACCAAGAA 108

Qy         24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
              |||:::   :::     |||:::     |||||||||||||||:::|||   |||
Db        109 GGTAGAAGTGAAAATTTAAAGAACAATCTCCTTGAAGGTAAGGAATTTGGTTTACAAGTA 168

Qy         44 GlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
              ||||||||||||||:::::::::|||   :::     |||
Db        169 GGTTTCCAGCGTTTCGTTCTTCTTGGTCAAATGTTAGGACTATGTGATGTT--------- 219

Qy         64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82
                 :::|||   :::|||      :::   ||||||      ::::::      :::
Db        220 ---CTGGATTCTTTCAACTTGAAGAATAAATCACTCGCAAAAAACGTCGCAATAATAAGA 276

Qy         83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
              ::::::::::::   |||         |||   ::::::   ||||||      ||||||
Db        277 GAACTAATCAGCACAATACAATTCAATAACGAAGAGGAAAATGTGGAAGGCCTTGAAAAA 336

Qy        103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
              :::  |||   :::||||||   |||:::|||         :::     :::
Db        337 AAATTGGTTAAACTTAAAAATAAGTATAGGTTAATATTGCTAGCTTTTCAAAGGGAATAT 396

Qy        123 Lys 123
              |||
Db        397 AAA 399

RESULT 5
CNS076JA/c
LOCUS       CNS076JA                1030 bp    DNA     linear   GSS 07-JUL-2001
DEFINITION  T7 end of clone BB0AA004B04 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL431468
VERSION     AL431468.1  GI:12214880
KEYWORDS    GSS.
SOURCE      Pichia angusta
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1030)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1030)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
   PUBMED   11152888
REFERENCE   3  (bases 1 to 1030)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1030
                     /organism="Pichia angusta"
                     /mol_type="genomic DNA"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA004B04"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    complement(<43. .>222)
                     /note="similar to Saccharomyces cerevisiae ORF YNL260c [
                     hypothetical protein ]"
                     /evidence=not_experimental
BASE COUNT      318 a    192 c    302 g    216 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              2.51e-12        Length:         1030
Score:                  187.00          Matches:        32
Percent Similarity:     76.67%          Conservative:   14
Best Local Similarity:  53.33%          Mismatches:     14
Query Match:            24.22%          Indels:         0
DB:                     29              Gaps:           0

US-09-980-054A-12 (1-148) x CNS076JA (1-1030)

Qy          3 AspIleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLys 22
              ||||||||   ||::::::|||||||||||   :::   |||:::   ||||||
Db        216 GACATAGACTTCGATTCGCTGTTGAACCTTGAGAACCAGTACTATCAGGAAGGTTTCATC 157

Qy         23 GluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGln 42
              |||||||||:::::::|||:::|||   |||:::|||||||||||:::   |||   |||
Db        156 GAAGGGCAGCTGGAGGGCTCGAAACAACAGTTTTTAGAAGGAAAGCAACTTGGCATCCAG 97

Qy         43 ThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeu 62
              ||||||||||||||   |||::::::|||       :::   |||:::   |||:::
Db         96 ACCGGTTTCCAGCGTCTTCTGGTCCTCGGTCAGTATAAAGCACTAGTCGCAATATGGATT 37

RESULT 6
CNS06PHM
LOCUS       CNS06PHM                849 bp    DNA     linear   GSS 05-JUL-2001
DEFINITION  T7 end of clone AV0AA014E08 of library AV0AA from strain CBS 379 of
            Saccharomyces exiguus, genomic survey sequence.
ACCESSION   AL409376
VERSION     AL409376.1  GI:12176838
KEYWORDS    GSS.
SOURCE      Saccharomyces exiguus
  ORGANISM  Saccharomyces exiguus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

```
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 849)
  AUTHORS   Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 6.
            Saccharomyces exiguus
  JOURNAL   FEBS Lett. 487 (1), 42-46 (2000)
  MEDLINE   20584716
   PUBMED   11152881
REFERENCE   3  (bases 1 to 849)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .849
                     /organism="Saccharomyces exiguus"
                     /mol_type="genomic DNA"
                     /strain="CBS 379"
                     /db_xref="taxon:34358"
                     /clone="AV0AA014E08"
                     /clone_lib="AV0AA"
                     /note="end : T7"
     misc_feature    <34. .>462
                     /note="similar to Saccharomyces cerevisiae ORF YNL260c [
                     hypothetical protein ]"
                     /evidence=not_experimental
BASE COUNT      311 a    108 c    159 g    253 t     18 others
ORIGIN

Alignment Scores:
Pred. No.:              1.68e-06        Length:         849
Score:                  138.50          Matches:        32
Percent Similarity:     53.10%          Conservative:   28
Best Local Similarity:  28.32%          Mismatches:     50
Query Match:            17.94%          Indels:         3
DB:                     29              Gaps:           2

US-09-980-054A-12 (1-148) x CNS06PHM (1-849)

Qy          22 LysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyr 41
               :::||||||              |||:::   :::|||   ||||||:::::::|||
Db          31 CGAGAGGGTCATAATKAAAATTTAAAGAATAACTTCCTCGAKGGGAAGCAATTTGGTTTA 90

Qy          42 GlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrp 61
               |||   |||||||||||||||:::::::::|||   :::      :::
Db          91 CAAGTTGGATTCCAAAGATTTGTTTTGTTAGGTCAAATGATTGGGATATGTKACGTG--- 147

Qy          62 LeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeu 81
               |||      :::|||      :::|||      :::|||      ::::::  ::::::

Db         148 CTAGATTCCTTAGATCTGWWGAGTAAT------GCATTAAGTAAAAATGTTCAAAGTGTA 201

Qy          82 GluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGlu 101
               ::::::  ::::::  :::||||||  |||:::  ||||||:::
Db         202 CGTTCTTTGATTAACGAGGTTGGCATGACAAATGATGATGAAAATGTAGAAGAATTAGWW 261

Qy         102 LysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThr 121
               |||  :::      :::||||||  |||  |||      :::
Db         262 AAGATAATGATTAAWCTGAAAAACAAATTTAGAACTATTATGATAACATTCCAAAGAATT 321

Qy         122 TrpLysIleAspSerLeuAspAsnLeuValLysGluVal 134
                  |||  |||      :::          ||||||
Db         322 ATGAAGAAGGATAAAAGAGAACCATTTACCTTTGAAGTT 360

RESULT 7
CB632715/c
LOCUS       CB632715                784 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSIIEb11C06.r OSIIEb Oryza sativa (indica cultivar-group) cDNA
            clone OSIIEb11C06 3', mRNA sequence.
ACCESSION   CB632715
VERSION     CB632715.1  GI:29627704
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 11  row: C  column: 06
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR36"
                     /db_xref="taxon:39946"
                     /clone="OSIIEb11C06"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSIIEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (PO6-6-3)"
BASE COUNT      228 a    186 c    136 g    234 t
ORIGIN

Alignment Scores:
Pred. No.:              1.66e-05        Length:         784
Score:                  130.00          Matches:        36
Percent Similarity:     52.14%          Conservative:   25
Best Local Similarity:  30.77%          Mismatches:     49
Query Match:            16.84%          Indels:         7
DB:                     14              Gaps:           3

US-09-980-054A-12 (1-148) x CB632715 (1-784)

Qy          12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
```

```
                ||||:::|||      |||:::   |||||||||   |||      :::|||
Db        593 TTAGATGAGACACACTATCAAACGGGTTTCAAGAATGGTTATAGTGAGGGCTTGGTGTCT 534

Qy         32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIle 51
                   ||||||::::::   |||   :::   |||||||||           :::
Db        533 GGAAAAGAAGAGGGAAGGCAGGTTGGTTTAAAGAATGGTTTTCAGGTAGGTGAAGAACTA 474

Qy         52 GlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------HisIleAspGlnTyrAsn 69
                ||:::   |||      :::      |||   |||          |||||||||
Db        473 GGTTTTTATCAGGGATGTCTGGATGTTTGGACGTCGTTGGTTTCAATTGATCAAGATGCA 414

Qy         70 AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSer 89
                ||:::   :::|||   ::::::      |||      ::::::
Db        413 TTCTCAGCTCGGGTCAGGAAAAACATTGAGCAACTAGCTGCACTGTTGAGAAGCTATCCG 354

Qy         90 IleThrAsn---GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArg 108
                ::::::|||      |||::::::|||:::|||         ::::::|||  |||
Db        353 TTGTCGAACCCAGAAGACGAGCAAGTTCAAGATATA---------ATGGAGAAGATAAGA 303

Qy        109 AsnLysLeuArgValIleAlaSer-IleThrLysGluThrTrpLysIle 124
                |||   |||||||||  ::: :::   :::   |||   |||
Db        302 CTGAAATTCAGGGTTATCACAGCAAGTTTAGGTACAAAACTGGAGTATC 254

RESULT 8
BQ461520
LOCUS       BQ461520                606 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HD04B05r HD Hordeum vulgare cDNA clone HD04B05 5-PRIME, mRNA
            sequence.
ACCESSION   BQ461520
VERSION     BQ461520.1  GI:21269291
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 606   Std Error: 0.00
            Plate: 4  row: B  column: 5
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Golden Promise"
                     /db_xref="taxon:4513"
                     /clone="HD04B05"
                     /tissue_type="callus"
                     /dev_stage="callus (5-10 mm in diameter)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HD"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable.Average insert size is 1
                     kb"
BASE COUNT      172 a    118 c    174 g    142 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.000113      Length:          606
Score:                    122.00        Matches:         32
Percent Similarity:       51.35%        Conservative:    25
Best Local Similarity:    28.83%        Mismatches:      50
Query Match:              15.80%        Indels:          4
DB:                       13            Gaps:            2

US-09-980-054A-12 (1-148) x BQ461520 (1-606)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              ::::::   ::::::   |||:::|||      |||:::   |||:::|||:::|||
Db        258 CTTGAACCAATGGTAGCCTTAGACGAGACACATTATCAAGATGGTTACAAAGATGGTTAT 317

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                 |||                 ||||||::::::   |||   :::   ||||||
Db        318 GATGATGGCTTGGTATCTGGGAAGGAAGAAGGAAGGCAGGTCGGTTTAAAGATGGGTTTC 377

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              |||           :::|||:::   |||      :::      |||:::|||   |||
Db        378 CAGGTAGGTGAAGAGCTAGGATTCTATCAGGGCTGCTTGGATGTGTGGATGTCAATAATT 437

Qy         66 ------AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                   ||||||         |||:::   :::|||   |||:::      |||
Db        438 TGCCTTGATCAAGATGCATTCTCAGCTCGGGTCAGGAAACACATGGAGCAATTAGCTGCA 497

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              :::::::::         ::::::      |||   |||           :::
Db        498 CTTCTAAGCAACTATCCTTTGTCT------GATCCAGAGAATAATCAGCTTCAAGACATG 551

Qy        104 IleLysLysAlaArgAsnLysLeuArgValIle 114
              :::|||      |||   |||   |||||||||
Db        552 ATGAAGGATATAAGGCTGAAATTCAGGGTTATC 584

RESULT 9
BU968430
LOCUS       BU968430                611 bp    mRNA    linear   EST 22-OCT-2002
DEFINITION  HB07H22r BC Hordeum vulgare subsp. vulgare cDNA clone HB07H22
            5-PRIME, mRNA sequence.
ACCESSION   BU968430
VERSION     BU968430.1  GI:24219223
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
  TITLE     Barley ESTs from developing seeds
  JOURNAL   Unpublished
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 611   Std Error: 0.00
            Plate: 7  row: H  column: 22
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="barke"
                     /db_xref="GABI:233254"
                     /db_xref="taxon:112509"
                     /clone="HB07H22"
                     /tissue_type="developing caryopsis"
                     /dev_stage="8-15 DAP (days after pollination)"
                     /lab_host="XL10-Gold"
                     /clone_lib="BC"
```

/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); developing caryopsis , 8-15 DAP(days after pollination) Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRIadapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinats is not 100% reliable."

BASE COUNT 171 a 119 c 178 g 143 t
ORIGIN

Alignment Scores:

Pred. No.:	0.000114	Length:	611
Score:	122.00	Matches:	32
Percent Similarity:	51.35%	Conservative:	25
Best Local Similarity:	28.83%	Mismatches:	50
Query Match:	15.80%	Indels:	4
DB:	13	Gaps:	2

US-09-980-054A-12 (1-148) x BU968430 (1-611)

```
Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              ::::::   ::::::   |||:::|||      |||:::   |||:::|||:::|||
Db        260 CTTGAACCAATGGTAGCCTTAGACGAGACACATTATCAAGATGGTTACAAAGATGGTTAT 319

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                |||                    ||||||::::::   |||   :::   ||||||
Db        320 GATGATGGCTTGGTATCTGGGAAGGAAGAAGGAAGGCAGGTCGGTTTAAAGATGGGTTTC 379

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              |||           :::|||:::   |||      :::      |||:::|||   |||
Db        380 CAGGTAGGTGAAGAGCTAGGATTCTATCAGGGCTGCTTGGATGTGTGGATGTCAATAATT 439

Qy         66 ------AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                 ||||||          |||:::   :::|||   |||:::       |||
Db        440 TGCCTTGATCAAGATGCATTCTCAGCTCGGGTCAGGAAACACATGGAGCAATTAGCTGCA 499

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              :::::::::        ::::::     |||   |||          :::
Db        500 CTTCTAAGCAACTATCCTTTGTCT------GATCCAGAGAATAATCAGCTTCAAGACATG 553

Qy        104 IleLysLysAlaArgAsnLysLeuArgValIle 114
              :::|||      |||   |||   |||||||||
Db        554 ATGAAGGATATAAGGCTGAAATTCAGGGTTATC 586
```

RESULT 10
BQ472468

LOCUS	BQ472468 641 bp mRNA linear EST 31-MAY-2002
DEFINITION	HB09H22T BC Hordeum vulgare subsp. vulgare cDNA clone HB09H22 5-PRIME, mRNA sequence.
ACCESSION	BQ472468
VERSION	BQ472468.1 GI:21284597
KEYWORDS	EST.
SOURCE	Hordeum vulgare subsp. vulgare
ORGANISM	Hordeum vulgare subsp. vulgare Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE	1 (bases 1 to 641)
AUTHORS	Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
TITLE	Barley ESTs from developing seeds
JOURNAL	Unpublished
COMMENT	Contact: Stein Nils Molecular Markers Group, Department Genbank Institute of Plant Genetics and Crop Plant Research (IPK) Corrensstr. 3, 06466, Gatersleben, Germany Tel: 039482-5522 Fax: 039482-5595 Email: stein@ipk-gatersleben.de Insert Length: 641 Std Error: 0.00 Plate: 9 row: H column: 22 Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .641
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/db_xref="taxon:112509"
/clone="HB09H22"
/tissue_type="developing caryopsis"
/dev_stage="8-15 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="BC"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); developing caryopsis , 8-15 DAP(days after pollination) Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRIadapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinats is not 100% reliable."

BASE COUNT 177 a 127 c 183 g 154 t
ORIGIN

Alignment Scores:

Pred. No.:	0.000121	Length:	641
Score:	122.00	Matches:	32
Percent Similarity:	51.35%	Conservative:	25
Best Local Similarity:	28.83%	Mismatches:	50
Query Match:	15.80%	Indels:	4
DB:	13	Gaps:	2

US-09-980-054A-12 (1-148) x BQ472468 (1-641)

```
Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              ::::::   ::::::   |||:::|||      |||:::   |||:::|||:::|||
Db        260 CTTGAACCAATGGTAGCCTTAGACGAGACACATTATCAAGATGGTTACAAAGATGGTTAT 319

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                |||                    ||||||::::::   |||   :::   ||||||
Db        320 GATGATGGCTTGGTATCTGGGAAGGAAGAAGGAAGGCAGGTCGGTTTAAAGATGGGTTTC 379

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
              |||           :::|||:::   |||      :::      |||:::|||   |||
Db        380 CAGGTAGGTGAAGAGCTAGGATTCTATCAGGGCTGCTTGGATGTGTGGATGTCAATAATT 439

Qy         66 ------AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                 ||||||          |||:::   :::|||   |||:::       |||
Db        440 TGCCTTGATCAAGATGCATTCTCAGCTCGGGTCAGGAAACACATGGAGCAATTAGCTGCA 499

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              :::::::::        ::::::     |||   |||          :::
Db        500 CTTCTAAGCAACTATCCTTTGTCT------GATCCAGAGAATAATCAGCTTCAAGACATG 553

Qy        104 IleLysLysAlaArgAsnLysLeuArgValIle 114
              :::|||      |||   |||   |||||||||
Db        554 ATGAAGGATATAAGGCTGAAATTCAGGGTTATC 586
```

RESULT 11
AI994649

LOCUS	AI994649 530 bp mRNA linear EST 08-SEP-1999
DEFINITION	701499129 A. thaliana, Ohio State clone set Arabidopsis thaliana cDNA clone 701499129, mRNA sequence.
ACCESSION	AI994649
VERSION	AI994649.1 GI:5841554
KEYWORDS	EST.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 530)
AUTHORS	Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,

```
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="701499129"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
BASE COUNT      127 a    118 c    115 g    170 t
ORIGIN

Alignment Scores:
Pred. No.:                0.000899        Length:          530
Score:                    114.00          Matches:         28
Percent Similarity:       48.11%          Conservative:    23
Best Local Similarity:    26.42%          Mismatches:      53
Query Match:              14.77%          Indels:          2
DB:                       9               Gaps:            1

US-09-980-054A-12 (1-148) x AI994649 (1-530)

Qy          6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
              :::|||   ::::::   |||||||||        :::   |||||||  |||||||
Db        155 CTCGATTGTATCGTGCGTTTAGAGGAGACACATGTTCAACAAGGCTTCGATGAGGGTTAC 214

Qy         26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
                ::|||                  :::   :::      |||   :::   ||||||
Db        215 GAAGAAGGTCTTGTGTCCGGTCGTGAAGACGCTCGTCATCTGGGTTTGAAACTCGGGTTC 274

Qy         46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------ 63
              :::         :::|||||:::   :::                |||   |||
Db        275 GAGACAGGCGAGCTGATTGGATTCTACAGAGGTTGCTCTGCTCTTTGGAATTCAGCTCTC 334

Qy         64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                ||||||              |||      |||      |||||||||:::
Db        335 CGTATTGATCCTACACGCTTCTCTCCTCAGCTCCATAAGCATCTCAATGATTTCCATGTC 394

Qy         84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
              ::::::   :::|||   :::   :::   |||   |||   :::|||      :::
Db        395 TTGCTCGATAAAATCCCGCTTTTGGATCCCGACGACGAAGCTAAAGACGGGATAAGGATG 454

Qy        104 IleLysLysAlaArgAsn 109
              |||   ::::::   :::
Db        455 ATCTCAGAGTCAAATTCA 472

RESULT 12
BZ992706/c
LOCUS       BZ992706                 879 bp    DNA     linear   GSS 25-MAR-2003
DEFINITION  PUEBB09TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa195B18,
            genomic survey sequence.
ACCESSION   BZ992706
VERSION     BZ992706.1  GI:29235764
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa195B18"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      274 a    203 c    168 g    234 t
ORIGIN

Alignment Scores:
Pred. No.:                0.00471         Length:          879
Score:                    110.50          Matches:         32
Percent Similarity:       48.74%          Conservative:    26
Best Local Similarity:    26.89%          Mismatches:      56
Query Match:              14.31%          Indels:          5
DB:                       29              Gaps:            3

US-09-980-054A-12 (1-148) x BZ992706 (1-879)

Qy          1 MetSerAspIleAsp---IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeu 19
              |||   ||   |||   ::::::      :::   |||:::|||      |||:::
Db        752 ATGATTGATGCTGATTTTCTTGAACCAACTGTACTCTTAGATGAGACACATTACCAGGAG 693

Qy         20 GlyPheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyr 39
              |||::::::   |||          |||                  ||||||::::::
Db        692 GGGTACAGAAATGGCTACAATGATGGTTTGGCATCTGGAAAGGAAGAGGGGAGGCAGGTT 633

Qy         40 GlyTyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLys 59
              |||   :::   |||||||||          :::|||:::          :::
Db        632 GGTTTAAAGATGGGTTTCCAAGTCGGTGAAGAACTGGGTTTCTTTCACGGCTGTCTGGAT 573

Qy         60 PheTrpLeuSerHis------IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHis 77
              |||   |||          |||||||||          ::::::   :::|||   :::
Db        572 GTGTGGACGTCTGCAATTCTCATTGATCAGAATGCCTTCACAGCTCGGGTCAGGAAGAAT 513

Qy         78 LeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluVal 97
              :::      |||      ::::::            ::::::      |||   |||
Db        512 ATTGAGCAACTGGCTGCACTGGTGGGCAGCTATCCGCTGTCC------GATCCGGAGAAT 459

Qy         98 GluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSer 116
              |||      :::      :::::::|||   |||   |||   |||||||||   :::
Db        458 GAACAAATTCAAGACGTGATGGAGAAGATAAGGCTGAAATTCAGGGTCATCACTGCA 402

RESULT 13
CA764105
LOCUS       CA764105                 477 bp    mRNA    linear   EST 08-JAN-2003
DEFINITION  AF53-Rpf_02_E02_T7_003.ab1 IRRI Drought Stress Panicle Library
            Oryza sativa (indica cultivar-group) cDNA clone C0000482 5' similar
            to unknown, mRNA sequence.
ACCESSION   CA764105
VERSION     CA764105.2  GI:27545938
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and
            Bruskiewich,R.M.
  TITLE     IRRI Drought Stress Panicle cDNA Library
  JOURNAL   Unpublished
COMMENT     On Dec 2, 2002 this sequence version replaced gi:25993360.
            Contact: Richard Bruskiewich
            Biometrics and Bioinformatics Unit
            International Rice Research Institute
            DAPO 7777, Metro Manila, Philippines
            Tel: +63-2-845-0563
            Fax: +63-2-845-0606
            Email: r.bruskiewich@cgiar.org
            International Rice Information System (IRIS;
            http://www.iris.irri.org): D0200481
            Assignment of putative function to the sequence by S. Rudd of the
            Munich Information Center for Protein Sequences (http://mips.gsf.de
            )
            Plate: 02  row: E  column: 02.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="C0000482"
                     /tissue_type="Panicles"
                     /dev_stage="Flowering"
                     /clone_lib="IRRI Drought Stress Panicle Library"
                     /note="Vector: pBluescript II SK+; Water stress was
                     applied by not watering for 4 consecutive days. Panicles
                     were collected from control (well watered) and stressed
                     plants at 2 days before heading, at heading, 50% flowering
                     and 4 days after 50% flowering."
BASE COUNT      143 a     69 c    126 g    139 t
ORIGIN

Alignment Scores:
Pred. No.:              0.00241        Length:         477
Score:                  110.00         Matches:        32
Percent Similarity:     52.29%         Conservative:   25
Best Local Similarity:  29.36%         Mismatches:     41
Query Match:            14.25%         Indels:         11
DB:                     14             Gaps:           4

US-09-980-054A-12 (1-148) x CA764105 (1-477)

Qy          20 GlyPheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyr 39
               |||:::   ||||||   :::   |||   :::            ||||||::::::
Db           2 GGTTATAGTGAGGGCTTGGTGTCTGGAAAAGAA-----------GAGGGAAGGCAGGTT 49

Qy          40 GlyTyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLys 59
               |||   :::   |||||||||           :::|||:::   |||       :::
Db          50 GGTTTAAAGAATGGTTTTCAGGTAGGTGAAGAACTAGGTTTTTATCAGGGATGTCTGGAT 109

Qy          60 PheTrpLeuSer------HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHis 77
                  |||   |||         |||||||||         |||:::   :::|||   :::
Db         110 GTTTGGACGTCGTTGGTTTCAATTGATCAAGATGCATTCTCAGCTCGGGTCAGGAAAAAC 169

Qy          78 LeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsn---GlyAspLysGlu 96
               :::      |||      ::::::           ::::::|||   |||   ::::::
Db         170 ATTGAGCAACTAGCTGCACTGTTGAGAAGCTATCCGTTGTCGAACCCAGGAGGCGAGCAA 229

Qy          97 ValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSer 116
               |||:::|||           :::::::|||   |||   |||   |||||||||   :::
Db         230 GTTCAAGATATA---------ATGGAGAAGATAAGACTGAAATTCAGGGTTATCACAGCA 280

Qy         117 -IleThrLysGluThrTrpLysIle 124
                :::   :::   |||   |||
Db         281 AGTTTAGGTACAAAACTGGAGTATC 305

RESULT 14
AV913233
```

```
LOCUS       AV913233                566 bp    mRNA    linear   EST 18-JAN-2002
DEFINITION  AV913233 K. Sato unpublished cDNA library, cv. Haruna Nijo
            germination shoots Hordeum vulgare subsp. vulgare cDNA clone
            bags21f17 5', mRNA sequence.
ACCESSION   AV913233
VERSION     AV913233.1  GI:18209010
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855 .
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="bags21f17"
                     /tissue_type="shoots"
                     /dev_stage="germination"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo germination shoots"
BASE COUNT      163 a    100 c    151 g    151 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              0.00303        Length:         566
Score:                  110.00         Matches:        29
Percent Similarity:     49.55%         Conservative:   26
Best Local Similarity:  26.13%         Mismatches:     52
Query Match:            14.25%         Indels:         4
DB:                     9              Gaps:           2

US-09-980-054A-12 (1-148) x AV913233 (1-566)

Qy           6 IleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
               ::::::      |||   |||:::|||      |||      |||::::::   |||
Db          97 CTTGAACCAGCATTAGCCTTAGATGAGACACATTATCACGAGGGTTATAGAAATGGGTAT 156

Qy          26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
               |||                  ||||||::::::   |||   :::   |||||||
Db         157 GATGATGGCCTGGTATCTGGAAAGGAAGAGGGGAGGCAGGTTGGCTTAAAGAATGGTTTC 216

Qy          46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------ 63
               |||            :::|||:::   |||       :::       |||:::|||
Db         217 CAGGTAGGCGAAGAACTGGGATTCTATCAGGGCTGCTTGGATGTGTGGATGTCAATAATT 276

Qy          64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
                  :::|||||           |||:::   :::|||   ::::::      |||
Db         277 CGCCTCGATCAAGATGCATTCTCAGCTCGTGTGAGGAAATACATAGAGCAACTAGCTGCA 336

Qy          84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
               :::::::::          ::::::      |||   |||   |||      :::
Db         337 CTTGTGAGCAACTATCCTTTGTCT------GATCCAGAGGATGAGCAGCTTCAAGGCATG 390

Qy         104 IleLysLysAlaArgAsnLysLeuArgValIle 114
               :::::::::   |||   |||      ||||||
Db         391 ATGAGGGAGATACGGCTGAAATTCAGCGTTATC 423

RESULT 15
CB632714
```

```
LOCUS       CB632714                 648 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSIIEb11C06.f OSIIEb Oryza sativa (indica cultivar-group) cDNA
            clone OSIIEb11C06 5', mRNA sequence.
ACCESSION   CB632714
VERSION     CB632714.1  GI:29627703
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 11  row: C  column: 06
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR36"
                     /db_xref="taxon:39946"
                     /clone="OSIIEb11C06"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSIIEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (PO6-6-3)"
BASE COUNT      155 a    150 c    182 g    161 t
ORIGIN
```

Alignment Scores:

Pred. No.:	0.00362	Length:	648
Score:	110.00	Matches:	26
Percent Similarity:	50.57%	Conservative:	18
Best Local Similarity:	29.89%	Mismatches:	41
Query Match:	14.25%	Indels:	2
DB:	14	Gaps:	1

US-09-980-054A-12 (1-148) x CB632714 (1-648)

```
Qy          12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
               |||:::|||      |||:::   |||||||||   |||      :::|||
Db         378 TTAGATGAGACACACTATCAAACGGGTTTCAAGAATGGTTATAGTGAGGGCTTGGTGTCT 437

Qy          32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIle 51
                     |||||:::::   |||   :::   |||||||||              :::
Db         438 GGAAAAGAAGAGGGAAGGCAGGTTGGTTTAAAGAATGGTTTTCAGGTAGGTGAAGAACTA 497

Qy          52 GlyTyrIleGlnGluLeuMetLysPheTrpLeuSer------HisIleAspGlnTyrAsn 69
               |||:::   |||       :::      |||   |||          |||||||||
Db         498 GGTTTTTATCAGGGATGTCTGGATGTTTGGACGTCGTTGGTTTCAATTGATCAAGATGCA 557

Qy          70 AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSer 89
                  |||:::   :::|||   ::::::      |||      ::::::
Db         558 TTCTCAGCTCGGGTCAGGAAAAACATTGAGCAACTAGCTGCACTGTTGAGAAGCTATCCG 617

Qy          90 IleThrAsnGlyAspLysGlu 96
                     ::::::|||   :::   |||
Db         618 TTGTCGAACCCAGAAGACGAG 638
```

Search completed: February 16, 2004, 16:07:51
Job time : 2539 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:10:37 ; Search time 35 Seconds
(without alignments)
1091.194 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	132	17.1	134	3	Q9Y7Q1	Q9y7q1 schizosacch
2	129	16.7	144	10	Q8GVP9	Q8gvp9 oryza sativ
3	115.5	15.0	139	5	Q8T1X1	Q8t1x1 dictyosteli
4	99.5	12.9	226	4	Q8WVZ5	Q8wvz5 homo sapien
5	99.5	12.9	226	4	Q9NRH1	Q9nrh1 homo sapien
6	99	12.8	222	2	Q9S0B1	Q9s0b1 borrelia bu
7	96	12.4	179	2	Q9S039	Q9s039 borrelia bu
8	96	12.4	223	16	Q8D3E3	Q8d3e3 wiggleswort
9	95.5	12.4	141	5	Q9V670	Q9v670 drosophila
10	95	12.3	1680	4	Q9P1Z9	Q9p1z9 homo sapien
11	94.5	12.2	192	2	Q9S062	Q9s062 borrelia bu
12	92	11.9	270	16	Q9KQ70	Q9kq70 vibrio chol
13	91.5	11.9	5251	5	Q8IID4	Q8iid4 plasmodium
14	91	11.8	239	16	Q8ZI06	Q8zi06 yersinia pe
15	91	11.8	247	16	Q8CKH8	Q8ckh8 yersinia pe
16	91	11.8	266	2	Q9Z6G6	Q9z6g6 vibrio para
17	90.5	11.7	746	5	Q25561	Q25561 naegleria f
18	90	11.7	178	2	Q44866	Q44866 borrelia bu
19	90	11.7	526	5	Q9VUA3	Q9vua3 drosophila
20	90	11.7	600	5	Q8IGY6	Q8igy6 drosophila
21	90	11.7	620	5	Q8SWR2	Q8swr2 drosophila
22	89.5	11.6	196	2	Q9S0H2	Q9s0h2 borrelia bu
23	89.5	11.6	707	5	Q8IBI2	Q8ibi2 plasmodium
24	89.5	11.6	964	5	Q8IDE1	Q8ide1 plasmodium
25	89	11.5	183	1	Q9C4P8	Q9c4p8 methanococc
26	89	11.5	210	2	Q45025	Q45025 borrelia bu
27	89	11.5	583	5	Q8ILQ1	Q8ilq1 plasmodium
28	89	11.5	1844	5	Q8I5U0	Q8i5u0 plasmodium
29	88.5	11.5	2020	5	Q8ID80	Q8id80 plasmodium
30	87.5	11.3	359	12	Q8V3S0	Q8v3s0 swinepox vi
31	87.5	11.3	527	11	Q8BVJ9	Q8bvj9 mus musculu
32	87.5	11.3	616	5	Q8I5Q2	Q8i5q2 plasmodium
33	87.5	11.3	755	4	Q9UFE4	Q9ufe4 homo sapien
34	87	11.3	190	2	Q9S078	Q9s078 borrelia bu
35	87	11.3	190	2	O07494	O07494 borrelia bu
36	87	11.3	263	16	O50958	O50958 borrelia bu
37	87	11.3	496	5	Q8IBU7	Q8ibu7 plasmodium
38	86.5	11.2	212	16	Q8YMI9	Q8ymi9 anabaena sp
39	86.5	11.2	409	5	Q9BLN9	Q9bln9 plasmodium
40	86.5	11.2	492	5	Q9BLN8	Q9bln8 plasmodium
41	86.5	11.2	876	11	Q99N77	Q99n77 rattus norv
42	86.5	11.2	930	11	Q99N78	Q99n78 rattus norv
43	86	11.1	2285	9	O64046	O64046 bacteriopha
44	86	11.1	2285	16	O31976	O31976 bacillus su
45	85.5	11.1	174	2	Q9L9M8	Q9l9m8 borrelia bu

ALIGNMENTS

RESULT 1
Q9Y7Q1
ID Q9Y7Q1 PRELIMINARY; PRT; 134 AA.
AC Q9Y7Q1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 15.5 kDa protein.
GN SPCC191.08.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Lyne M., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL049644; CAB41054.1; -.
DR GeneDB_SPombe; SPCC191.08; -.
KW Hypothetical protein.
SQ SEQUENCE 134 AA; 15485 MW; D3CE4F6135531AF0 CRC64;

Query Match 17.1%; Score 132; DB 3; Length 134;
Best Local Similarity 28.0%; Pred. No. 0.0041;
Matches 37; Conservative 29; Mismatches 58; Indels 8; Gaps 5;

```
Qy          4 IDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS 63
              :| : |  ||| :|: |: || ::| :  | |   :| :  : ::|: | |   : |||
Db          1 MDWEEVTKLEENEYKRGYDEGILKGIEQGYEEAFLFGLEHAYNKYLLAGEIYGRVCFWLK 60

Qy         64 HID-QYNNSSSLRNHLNNLEDIMAQISITNGDKEVED--YEKNIKKARNKLRVIASI--T 118
                : |:          ||  |: ::  :  || : |  |  ::     |   | :|::|:  |
Db         61 EENSQHPKIKKAHRHLEQLKSLLESLP-TNNELEETDAGFDSYWNKITAKAKVVSSLLGT 119

Qy        119 K--ETWKIDSLD 128
               |     |||: |
Db        120 KILPAEKIDAND 131
```

```
RESULT 2
Q8GVP9
ID   Q8GVP9      PRELIMINARY;      PRT;   144 AA.
AC   Q8GVP9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   P0681F05.15 protein.
GN   P0681F05.15.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
RT   clone:P0681F05.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP004674; BAC45146.1; -.
SQ   SEQUENCE   144 AA;  16311 MW;  DF840964D704F90B CRC64;

  Query Match             16.7%;  Score 129;  DB 10;  Length 144;
  Best Local Similarity   31.5%;  Pred. No. 0.0076;
  Matches   34;  Conservative   23;  Mismatches   45;  Indels    6;  Gaps    3;

Qy         12 LEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS--HIDQYN 69
              |:|   |: ||| |  :|       ||:: | : |||    :|: |  :  | |   |||
Db         18 LDETHYQTGFKNGYSEGLVSGKEEGRQVGLKNGFQVGEELGFYQGCLDVWTSLVSIDQDA 77

Qy         70 NSSSLRNHLNNLEDIMAQISITN-GDKEVEDYEKNIKKARNKLRVIAS 116
              |: :| ::  |  ::    ::|  |::|:|    ::| | | ||| :
Db         78 FSARVRKNIEQLAALLRSYPLSNPEDEQVQDI---MEKIRLKFRVITA 122


RESULT 3
Q8T1X1
ID   Q8T1X1      PRELIMINARY;      PRT;   139 AA.
AC   Q8T1X1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 16.0 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116032; AAL93047.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   139 AA;  16031 MW;  A2BD240D720FDFA1 CRC64;

  Query Match             15.0%;  Score 115.5;  DB 5;  Length 139;
  Best Local Similarity   24.8%;  Pred. No. 0.084;
  Matches   33;  Conservative   35;  Mismatches   56;  Indels    9;  Gaps    4;

Qy          5 DIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              : | :|::| : |    ::|   | :  |:|| : |:: | :    ||| |  :  | :|
Db          3 EFDQLLSVESDAYISSKEQGIDDGKRLGYVEGYQLGFEKGIELGQEIGYYQSCVTVW-NH 61

Qy         65 IDQY--NNSSSLRNHLNNLEDIMAQI-SITNGDKEVEDY-----EKNIKKARNKLRVIAS 116
              :     ||::  |: |||:  :  | ::    | :|||     ::||    :::|:
Db         62 LVSINNNNNNNNNNNKNNLKFSVRGIQNLEKLTKLLEDYHLDFNDENIMNTLSEIRLKFK 121

Qy        117 ITKETWKIDSLDN 129
                :|     : : :|
Db        122 LTSVQLGLQTKEN 134


RESULT 4
Q8WVZ5
ID   Q8WVZ5      PRELIMINARY;      PRT;   226 AA.
AC   Q8WVZ5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC022043; AAH22043.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   226 AA;  25313 MW;  8A40E87864E3F14C CRC64;

  Query Match             12.9%;  Score 99.5;  DB 4;  Length 226;
  Best Local Similarity   26.7%;  Pred. No. 2.7;
  Matches   32;  Conservative   23;  Mismatches   46;  Indels   19;  Gaps    3;

Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
              |::    | |:::|     |        :|   ||: | :  |  | ::  :   ||    :||
Db         37 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 95

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              :|:| | :||| | : |           |:|           |: : |||  :  :| ||::
Db         96 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 137


RESULT 5
Q9NRH1
ID   Q9NRH1      PRELIMINARY;      PRT;   226 AA.
AC   Q9NRH1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GK003.
GN   GK003.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver cancer;
RA   Li Y., Wu T., Xu S., Ren S., Chen Z., Han Z.;
RT   "A novel gene expressed in human liver cancer tissue.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF226046; AAF86946.1; -.
SQ   SEQUENCE   226 AA;  25299 MW;  403133C215381E79 CRC64;

  Query Match             12.9%;  Score 99.5;  DB 4;  Length 226;
  Best Local Similarity   26.7%;  Pred. No. 2.7;
  Matches   32;  Conservative   23;  Mismatches   46;  Indels   19;  Gaps    3;

Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNN 70
              |::    | |:::|     |        :|   ||: | :  |  | ::  :   ||    :||
Db         37 NMQRRVKE-GYRDGIDAGKAVTLQQGFNQGYKKGAEVILNYGRLRGTLSALLSWCHLHNN 95

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              :|:| | :||| | : |           |:|           |: : |||  :  :| ||::
Db         96 NSTLINKINNLLDAVGQ---------CEEY---------VLKHLKSITPPSHVVDLLDSI 137
```

```
RESULT 6
Q9S0B1
ID   Q9S0B1      PRELIMINARY;      PRT;   222 AA.
AC   Q9S0B1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Conserved hypothetical protein.
GN   BBM34.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid cp32-6.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,
RA   White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van-Vugt R.,
RA   Palmer N., Haft D., Rosa P., Stevenson B.;
RT   "A bacterial genome in flux: The twelve linear and nine circular
RT   extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT   spirochete Borrelia burgdorferi.";
RL   Mol. Microbiol. 0:0-0(1999).
DR   EMBL; AE001578; AAF07545.1; -.
DR   InterPro; IPR003900; KID_repeat.
DR   Pfam; PF02524; KID; 9.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   222 AA;  25444 MW;  1E7F5A0EAA303BDE CRC64;

  Query Match             12.8%;  Score 99;  DB 2;  Length 222;
  Best Local Similarity   30.2%;  Pred. No. 2.8;
  Matches   32;  Conservative   22;  Mismatches   48;  Indels    4;  Gaps    4;

Qy         37 KEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKE 96
              | |  ::    || ||: :| : |   | | | |   |:  : ::|  : |  |:| | :
Db          7 KTYNIESIKNEFLNIGFSKEAIDFVFLHNDNY-NFEFLKEKIIDIEKNL-QKDISNLDAK 64

Qy         97 VEDYEKNIKKARNKLRV-IASITKE-TWKIDSLDNLVKEVGGTLQV 140
              ::: |||:    : | : | :: |    ||||||  :  |    |:
Db         65 IDNVEKNLNLKIDGLNIKIDNVEKSLNAKIDSLDTKIDNVEKNLNI 110


RESULT 7
Q9S039
ID   Q9S039      PRELIMINARY;      PRT;   179 AA.
AC   Q9S039;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Conserved hypothetical protein.
GN   BBN34.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid cp32-9.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,
RA   White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van-Vugt R.,
RA   Palmer N., Haft D., Rosa P., Stevenson B.;
RT   "A bacterial genome in flux: The twelve linear and nine circular
RT   extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT   spirochete Borrelia burgdorferi.";
RL   Mol. Microbiol. 0:0-0(1999).
DR   EMBL; AE001581; AAF07675.1; -.
DR   InterPro; IPR003900; KID_repeat.
DR   Pfam; PF02524; KID; 7.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   179 AA;  20713 MW;  40C1D64732A5CC8A CRC64;

  Query Match             12.4%;  Score 96;  DB 2;  Length 179;
  Best Local Similarity   31.2%;  Pred. No. 3.8;
  Matches   34;  Conservative   20;  Mismatches   47;  Indels    8;  Gaps    5;

Qy         37 KEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKE 96
              : |  ::    || ||: :| : |   | | | |   |:  : ::|  : |  |:: | :
Db          7 RTYDIESIKNEFLNIGFSEEAIDFVFLHNDNY-NYEVLKEKIIDVEKNL-QKDISSLDTK 64

Qy         97 VEDYEKN----IKKARNKLRV-IASITKE-TWKIDSLDNLVKEVGGTLQ 139
              ::: |||    |    | : | |:  |   |||||  :  |  |||
Db         65 IDNVEKNLNVKIDNVEKNLNIKIDSVKNELNSKIDSLDTKIDNVEKTLQ 113


RESULT 8
Q8D3E3
ID   Q8D3E3      PRELIMINARY;      PRT;   223 AA.
AC   Q8D3E3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   FliH protein.
GN   FLIH.
OS   Wigglesworthia brevipalpis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Wigglesworthia.
OX   NCBI_TaxID=164609;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22297718; PubMed=12219091;
RA   Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA   Aksoy S.;
RT   "Genome sequence of the endocellular obligate symbiont of tsetse
RT   flies, Wigglesworthia glossinidia.";
RL   Nat. Genet. 32:402-407(2002).
DR   EMBL; AB063521; BAC24204.1; -.
KW   Complete proteome.
SQ   SEQUENCE   223 AA;  25650 MW;  CC5697B9884937B0 CRC64;

  Query Match             12.4%;  Score 96;  DB 16;  Length 223;
  Best Local Similarity   27.6%;  Pred. No. 4.9;
  Matches   40;  Conservative   27;  Mismatches   46;  Indels   32;  Gaps    7;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHID 66
              |  || |    ::||: |         | |||| ||: ||:: ||    |:   |  :  |:
Db         48 DENLNYE----KIGFESG--------YKEGKEIGYKKGFEKGLI--EIKNEKKKKILQIE 93

Qy         67 QYNNSSSLRNHLNNLEDI----MAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
              ::     |:: |  |: :    : ||:|   :| ::          :: |     | |||
Db         94 KF--LVELKSSLIGLDSVIPSKIMQIAIKISEKIIK---------QSPLCNTKVILKETQ 142

Qy        123 KIDSLDNLVKEVGGTLQVSENPDDM 147
              |:    |    :|  | :  :|||:
Db        143 KLLHNKNF---LGNNLSIYVHPDDI 164


RESULT 9
Q9V670
ID   Q9V670      PRELIMINARY;      PRT;   141 AA.
AC   Q9V670; Q9V671;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG13175 protein.
GN   CG13175 OR CG13174.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
```

```
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003823; AAF58560.2; -.
DR   FlyBase; FBgn0033693; CG13175.
SQ   SEQUENCE   141 AA;  15724 MW;  CA0016D1658BAC84 CRC64;

  Query Match            12.4%;  Score 95.5;  DB 5;  Length 141;
  Best Local Similarity  25.5%;  Pred. No. 3.2;
  Matches   28;  Conservative   29;  Mismatches   42;  Indels   11;  Gaps    5;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHI- 65
              |:::  ||::  ||::|     | ||   :| | ||: |: : : :|  :|| |  |  :
Db         13 DDIVLTEEKEARLGYEE----GLKDGQEQGNEEGYKLGYAQGVSLG--EELGKI-LGQVV 65

Qy         66 --DQYNNSSSLRNHLNNLEDIMAQISITNGDK-EVEDYEKNIKKARNKLR 112
                |  ::  :|  |  |  :: :   ||  : ::    ::|: :  :||
Db         66 AQQQLKHTDKVRRSLEQLRSLIEEFPRTNDPQADIVGAVQDIRSSHRRLR 115


RESULT 10
Q9P1Z9
ID   Q9P1Z9      PRELIMINARY;      PRT;  1680 AA.
AC   Q9P1Z9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA1529 (Fragment).
GN   KIAA1529.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human
RT   genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT   which code for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
DR   EMBL; AB040962; BAA96053.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   1680 AA;  194698 MW;  BE3AF4E4E0013288 CRC64;

  Query Match            12.3%;  Score 95;  DB 4;  Length 1680;
  Best Local Similarity  25.2%;  Pred. No. 60;
  Matches   36;  Conservative   27;  Mismatches   40;  Indels   40;  Gaps    7;

Qy         24 GQIQGTKDQYLEGKEYGY-----QTGFQRFLIIGYIQELMKFWLSH---------- 64
              | :||  :: ||  :  :     || :|   :  | ||::: | :|
Db        660 GALQGKVEEDLELLDKSFETLADQTEWQSSHLFKYFQEVVQLWEAHQSELLVQELELEKR 719

Qy         65 IDQYNNSSSLRN-----HLNNLEDIMAQISITNGDKE--------VEDYEKNIKKARNKL 111
              ::|:    || :      ||: | | : | |    |||         |:|| ||:|
Db        720 MEQHRQKHSLESQVQEAHLDRLLDQLRQQS----DKETLAFHLEKVKDYLKNMKSRYECF 775

Qy        112 RVIASITKETWKIDSLDNLVKEV 134
                :  :|||  :  ::  ::||:
Db        776 HTL--LTKEVMEYPAI--MLKEL 794


RESULT 11
Q9S062
ID   Q9S062      PRELIMINARY;      PRT;   192 AA.
AC   Q9S062;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Conserved hypothetical protein.
GN   BBL35.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OG   Plasmid cp32-8.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,
RA   White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van-Vugt R.,
RA   Palmer N., Haft D., Rosa P., Stevenson B.;
RT   "A bacterial genome in flux: The twelve linear and nine circular
RT   extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT   spirochete Borrelia burgdorferi.";
RL   Mol. Microbiol. 0:0-0(1999).
DR   EMBL; AE001580; AAF07620.1; -.
DR   InterPro; IPR003900; KID_repeat.
DR   Pfam; PF02524; KID; 7.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   192 AA;  22050 MW;  B419C48522D0180E CRC64;

  Query Match             12.2%;  Score 94.5;  DB 2;  Length 192;
  Best Local Similarity   29.0%;  Pred. No. 5.4;
  Matches   40;  Conservative   29;  Mismatches   54;  Indels   15;  Gaps    8;

Qy          8 NVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHID- 66
              |: |:: |  |:|| |  |      :|    | ::  | :  :|   : | |  :|::|
Db         10 NIGNIKNEFLEIGFSEEAIDFV---FLHNDNYNFE--FLKEKLINLEKNLQKD-ISNLDI 63

Qy         67 QYNNSSSLRNHLNNLEDIM---AQISITNGDKEVEDYEKNIKKARNKLRV-IASITKETW 122
              : ||   ::| ||   | :    |  |:: | :::  |||::|  : |   | |: |
Db         64 KINN---VKNELNAKIDSVEKNLQKDISSLDIKIDSVEKNLQKDISSLNTKIDSVEKNLQ 120

Qy        123 K-IDSLDNLVKEVGGTLQ 139
               | | ||:  :  |  :||
Db        121 KDISSLNTKIDSVEKSLQ 138


RESULT 12
Q9KQ70
ID   Q9KQ70      PRELIMINARY;      PRT;   270 AA.
AC   Q9KQ70;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Flagellar assembly protein FliH, putative.
GN   VC2131.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004286; AAF95276.1; -.
DR   TIGR; VC2131; -.
DR   InterPro; IPR000563; Flag_FliH.
DR   Pfam; PF02108; FliH; 1.
DR   PRINTS; PR01003; FLGFLIH.
KW   Complete proteome.
SQ   SEQUENCE   270 AA;  29994 MW;  4934AAEFB8939C4F CRC64;

  Query Match             11.9%;  Score 92;  DB 16;  Length 270;
  Best Local Similarity   22.3%;  Pred. No. 13;
  Matches   33;  Conservative   31;  Mismatches   42;  Indels   42;  Gaps    6;

Qy         13 EEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSS 72
              : | |: ||::|: :| :  : ||:  ||| |      :  ||| :| ::: :|:
Db         88 QAEGYQQGFEQGKAEGFQAGHQEGQTQGYQDGVAEGQAL--IQEQVKTFMALANQFAQPL 145

Qy         73 SLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKID------- 125
              | |         ||:     :|:: |              :: ::|||   ::
Db        146 DLLN---------AQV-----EKQLVD------------MVLALTKEVVHVEVQTNPQV 178

Qy        126 SLDNLVKEV------GGTLQVSENPDDM 147
              || :   |      |  : :  ||:|:
Db        179 ILDTVKASVEALPIAGHAITLKLNPEDV 206


RESULT 13
Q8IID4
ID   Q8IID4      PRELIMINARY;      PRT;  5251 AA.
AC   Q8IID4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dynein heavy chain, putative.
GN   PF11_0240.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014839; AAN35824.1; -.
SQ   SEQUENCE   5251 AA;  617371 MW;  14C220359A132AD1 CRC64;

  Query Match             11.9%;  Score 91.5;  DB 5;  Length 5251;
  Best Local Similarity   23.0%;  Pred. No. 4.2e+02;
  Matches   41;  Conservative   35;  Mismatches   39;  Indels   63;  Gaps    9;

Qy          5 DIDNVLNLEEEQYE---------------LGFKEGQIQGTKDQYLEGKEYGYQTGFQRF 48
              | |||: | ::|: |               | || :          |:||    |
Db       1587 DDDNIKNDKDEKDEKEETIIKLISSYGEEICNFHEGLV-------LKGKVECYLND---- 1635

Qy         49 LIIGYIQELMKFWLSHI----DQYNNSSSL---RNHL-------------NNLEDIMAQI 88
              || :|:  :|::::::    | :||       |:|                |::|:|: :
Db       1636 -IIDHIKYTLKYYITNLFRLKDLFNNEKEKWIDENYLAQVFILCNTIFFVNDVENILIKK 1694

Qy         89 SITNGDKEVEDYEKN--------IKKA------RNKLRVIASITKETWKIDSLDNLVK 132
              | |  :|:  | ||         |||        :|:::::  || :|:  | |: ::|
Db       1695 DI-NIYEELNKYYKNHILQLENVIKKVQKKLTIKNRIKIMCIITLDTFYRDVLEVILK 1751


RESULT 14
Q8ZI06
ID   Q8ZI06      PRELIMINARY;      PRT;   239 AA.
AC   Q8ZI06;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative flagellar assembly protein.
GN   FLIH OR YPO0716.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414144; CAC89567.1; -.
DR   InterPro; IPR000563; Flag_FliH.
DR   Pfam; PF02108; FliH; 1.
DR   PRINTS; PR01003; FLGFLIH.
KW   Flagella; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   239 AA;  26725 MW;  6088795DBE257B09 CRC64;

  Query Match             11.8%;  Score 91;  DB 16;  Length 239;
  Best Local Similarity   26.5%;  Pred. No. 13;
  Matches   41;  Conservative   25;  Mismatches   57;  Indels   32;  Gaps    7;

Qy          7 DNVLNLEEEQYEL--GFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              |  |:  | | :|  ||:||  ||       ||||  ||| | :    :|:   | |
Db         35 DQTLDPAEYQKQLMAGFQEGISQGFDKGLAEGKEEGYQEGVR----LGHDDGLKK----- 85

Qy         65 IDQYNNSSSLRNHLNNLEDIMAQIS--ITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
                      :: | :  |::   |  ||     :| ||:  ::    |::: :|::
Db         86 ----GRIEGRQSELASFNDVIKPFSGYITQLHTYLETYEQ--RRRDELLQLVEKVTRQVI 139

Qy        123 KID------SLDNLVKE-------VGGTLQVSENP 144
               : :       |  ||:|       |   |:|  ||
Db        140 RCELALQPAQLLTLVEEALAALPMVPQQLKVYLNP 174


RESULT 15
Q8CKH8
ID   Q8CKH8      PRELIMINARY;      PRT;   247 AA.
AC   Q8CKH8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical.
GN   Y3462.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
DR   EMBL; AE013949; AAM87011.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   247 AA;  27690 MW;  666022F0B496D008 CRC64;

  Query Match             11.8%;  Score 91;  DB 16;  Length 247;
  Best Local Similarity   26.5%;  Pred. No. 14;
  Matches   41;  Conservative   25;  Mismatches   57;  Indels   32;  Gaps    7;

Qy          7 DNVLNLEEEQYEL--GFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSH 64
              |  |:  | | :|  ||:||  ||       ||||  ||| | :    :|:   | |
Db         43 DQTLDPAEYQKQLMAGFQEGISQGFDKGLAEGKEEGYQEGVR----LGHDDGLKK----- 93

Qy         65 IDQYNNSSSLRNHLNNLEDIMAQIS--ITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
                      :: | :  |::   |  ||     :| ||:  ::    |::: :|::
Db         94 ----GRIEGRQSELASFNDVIKPFSGYITQLHTYLETYEQ--RRRDELLQLVEKVTRQVI 147

Qy        123 KID------SLDNLVKE-------VGGTLQVSENP 144
               : :       |  ||:|       |   |:|  ||
Db        148 RCELALQPAQLLTLVEEALAALPMVPQQLKVYLNP 182


Search completed: February 16, 2004, 14:14:12
Job time : 35 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:09:06 ; Search time 12 Seconds
(without alignments)
579.996 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDNVLNLEEEQYELG.....NLVKEVGGTLQVSENPDDMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	223	28.9	198	1	YN00_YEAST	P53846	saccharomyc
2	97.5	12.6	610	1	MUTL_BORBU	O51229	borrelia bu
3	94	12.2	627	1	FLGK_BORBU	P70859	borrelia bu
4	91.5	11.9	1005	1	RA50_METJA	Q58718	methanococc
5	87	11.3	1539	1	Y373_HUMAN	O15078	homo sapien
6	84.5	10.9	852	1	RA50_THEMA	Q9x1x1	thermotoga
7	83.5	10.8	879	1	RA50_PYRHO	O58687	pyrococcus
8	83.5	10.8	880	1	RA50_PYRAB	Q9uzc8	pyrococcus
9	83.5	10.8	930	1	TRP6_MOUSE	Q61143	mus musculu
10	82.5	10.7	1225	1	SMC1_YEAST	P32908	saccharomyc
11	82	10.6	1966	1	MYSB_CAEEL	P02566	caenorhabdi
12	81	10.5	320	1	K6PF_SALTY	Q8xg19	salmonella
13	81	10.5	521	1	SYK_BORBU	O51603	borrelia bu
14	81	10.5	564	1	M12_STRPY	P19401	streptococc
15	81	10.5	1453	1	Y373_BOVIN	Q9tu23	bos taurus
16	81	10.5	2418	1	SPCA_HUMAN	P02549	homo sapien
17	80	10.4	978	1	RA50_AQUAE	O67124	aquifex aeo
18	79.5	10.3	1699	1	DPOL_THEG8	Q9hh84	thermococcu
19	79	10.2	1173	1	ATC2_YEAST	P38929	saccharomyc
20	78.5	10.2	385	1	YJ42_YEAST	P47122	saccharomyc
21	78.5	10.2	472	1	XISA_ANASP	P08862	anabaena sp
22	78.5	10.2	704	1	MEPB_RAT	P28826	rattus norv
23	78.5	10.2	714	1	YJL2_YEAST	P47025	saccharomyc
24	78.5	10.2	903	1	YB56_METJA	Q58556	methanococc
25	78.5	10.2	956	1	SYI_AQUAE	O66651	aquifex aeo
26	78.5	10.2	2245	1	MYSJ_DICDI	P54697	dictyosteli
27	78.5	10.2	4486	1	DYH9_HUMAN	Q9nyc9	homo sapien
28	78	10.1	382	1	SPCN_MOUSE	P16546	mus musculu
29	78	10.1	460	1	HEMN_PSEAE	P77915	pseudomonas
30	78	10.1	638	1	YAB1_YEAST	P31376	saccharomyc
31	78	10.1	1221	1	YMP3_CAEEL	Q10947	caenorhabdi
32	78	10.1	1839	1	CYAA_SACKL	P23466	saccharomyc
33	77.5	10.0	534	1	IF2P_SULAC	P95691	sulfolobus
34	77	10.0	228	1	FLIH_ECOLI	P31068	escherichia
35	77	10.0	237	1	GIDB_MYCPE	Q8euw9	mycoplasma
36	77	10.0	320	1	K6P1_ECOLI	P06998	escherichia
37	77	10.0	695	1	YCX7_CHLVU	O20159	chlorella v
38	77	10.0	1607	1	LMG1_MOUSE	P02468	mus musculu
39	76.5	9.9	332	1	SD22_SCHPO	P22194	schizosacch
40	76.5	9.9	496	1	ARAA_BACSU	P94523	bacillus su
41	76.5	9.9	734	1	KDGA_PIG	P20192	sus scrofa
42	76.5	9.9	864	1	RA50_SULSO	Q97wh0	sulfolobus
43	76.5	9.9	3587	1	SRF2_BACSU	Q04747	bacillus su
44	76	9.8	862	1	MUTS_BORBU	O51737	borrelia bu
45	76	9.8	1163	1	SBCC_CLOAB	Q97fk1	clostridium

ALIGNMENTS

RESULT 1
YN00_YEAST
ID YN00_YEAST STANDARD; PRT; 198 AA.
AC P53846;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 23.3 kDa protein in ORC5-ATX1 intergenic region.
GN YNL260C OR N0838.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldener U., Beinhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X96722; CAA65484.1; -.
DR EMBL; Z71536; CAA96167.1; -.
DR PIR; S63233; S63233.
DR SGD; S0005204; YNL260C.
KW Hypothetical protein.
SQ SEQUENCE 198 AA; 23328 MW; 2560D990E8A709F1 CRC64;

Query Match 28.9%; Score 223; DB 1; Length 198;
Best Local Similarity 35.9%; Pred. No. 3.7e-11;
Matches 47; Conservative 30; Mismatches 40; Indels 14; Gaps 3;

```
Qy          4 IDIDNVLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLS 63
              :| ||:|||||: |: || ||| :  |  :||||:|| | ||||| ::| :: |
Db         37 MDFDNLLNLEEQYYQEGFLEGQNENIKQSFLEGKQYGLQVGFQRFTLLGQMEGLCDV--- 93

Qy         64 HIDQYN-NSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETW 122
               |:|  :| :|  ::: :  :| : : | |: | ::|: : | :|| | |
Db         94 -IESYGLHSPTLEKNIHTIRTLMKGLKMNNDDESVMEFERVLIKLKNKFRTI-------- 144

Qy        123 KIDSLDNLVKE 133
               : :|  |||:
Db        145 -LITLHRLVKD 154
```

RESULT 2

```
MUTL_BORBU
ID   MUTL_BORBU     STANDARD;      PRT;   610 AA.
AC   O51229;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA mismatch repair protein mutL.
GN   MUTL OR BB0211.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
CC       DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
CC       REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
CC       PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
CC       DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
CC       BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001131; AAC66594.1; -.
DR   PIR; C70126; C70126.
DR   HSSP; P23367; 1BKN.
DR   TIGR; BB0211; -.
DR   HAMAP; MF_00149; -; 1.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR002099; DNA_mis_repair.
DR   Pfam; PF01119; DNA_mis_repair; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   TIGRFAMs; TIGR00585; mutl; 1.
DR   PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW   DNA repair; Complete proteome.
SQ   SEQUENCE   610 AA;  71542 MW;  36A41A8A884EC9FA CRC64;

  Query Match             12.6%;  Score 97.5;  DB 1;  Length 610;
  Best Local Similarity   24.7%;  Pred. No. 1.4;
  Matches   37;  Conservative   27;  Mismatches   51;  Indels   35;  Gaps    6;

Qy          1 MSDIDIDNVLNLEE-----EQYELGFKE-GQIQGTKDQYLEGKEYGYQTGFQRFLIIGYI 54
              : :|  || | ||      |: |: |  |||                   | |||:  |
Db        407 IGNIFSDNFLEFEEPPNKNKEKEEIKFNYIGQI-----------------FSEFLIVEKI 448

Qy         55 QELMKFWLSHIDQYNNS-----SSLRNHLNNLEDIMAQISITNGDKEVED-YEKNIKKAR 108
               |:      |||:          |||   |:: ::  |  |  || :|:  :  |:: :
Db        449 NEIY-----FIDQHAVHEKIIYEKLRNSKKNVQKLLVPIEFTVVDKNIEEIIDSEIEEYK 503

Qy        109 NKLRVIASITKETWKIDSLDNLVKEVGGTL 138
              :|: |  : ::::|: |:  :   ||
Db        504 QMDIIISKIGPKKYQLESIPNICSQYENTL 533
```

```
RESULT 3
FLGK_BORBU
ID   FLGK_BORBU     STANDARD;      PRT;   627 AA.
AC   P70859; O51199;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar hook-associated protein 1 (HAP1).
GN   FLGK OR BB0181.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=212;
RA   Ge Y., Old I.G., Saint-Girons I., Charon N.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U62901; AAB62740.1; -.
DR   EMBL; AE001129; AAC66579.1; -.
DR   EMBL; AJ003222; CAA06002.1; -.
DR   PIR; E70122; E70122.
DR   TIGR; BB0181; -.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   InterPro; IPR002371; Flag_hookAP1.
DR   PRINTS; PR01005; FLGHOOKAP1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; FALSE_NEG.
KW   Flagella; Complete proteome.
FT   CONFLICT    330    330       F -> S (IN REF. 1).
FT   CONFLICT    551    551       A -> R (IN REF. 1).
FT   CONFLICT    554    554       A -> R (IN REF. 1).
FT   CONFLICT    558    558       A -> R (IN REF. 1).
FT   CONFLICT    607    627       ASKFITVSVELIDTVINKMGV -> QVNLSLFLLN (IN
FT                                REF. 1).
SQ   SEQUENCE   627 AA;  70574 MW;  6D38B2F934900FF1 CRC64;

  Query Match             12.2%;  Score 94;  DB 1;  Length 627;
  Best Local Similarity   25.7%;  Pred. No. 2.7;
  Matches   49;  Conservative   23;  Mismatches   51;  Indels   68;  Gaps   10;

Qy         10 LNLEEEQYELGFKEG----QIQGTKDQYL------EGKEYGYQTGFQRFLII-------- 51
              ||  ::| :||  :|      |    ||: |      |    || |   :|: |
Db         55 LNRAKKQGQLG--QGIVVQSIDRVKDELLNTRIIEESHRLGYWTSQDKFISILEDVYNEP 112

Qy         52 ---GYIQELMKFWLSHIDQYNNSSSL----------------RNHLNNLED--IMA--Q 87
                  : |  || |  |  |     |                  ||  ::||   |||  :
Db        113 EDQSIRKRLNDFWESWHDLANQPQGLAERKIILERGKSPCEGIRNRFHSLERIYIMANDE 172
```

```
Qy          88 ISITNGDKEVEDYEKNIKKARNKLRVIASITKETWK----------IDSLDNLVKEVGG 136
               | ||  :           | | :| ||:: |:  |          :|: | :|:::|
Db         173 IKITTDE----------ANNYIRNIANLNKQISKSQAMKDNPNDLMDARDLMVEKLGN 220

Qy         137 TLQVS-ENPDD 146
                : || ||  |
Db         221 IISVSIENKQD 231
```

RESULT 4
RA50_METJA
ID RA50_METJA STANDARD; PRT; 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U67572; AAB99331.1; -.
DR PIR; A64465; A64465.
DR TIGR; MJ1322; -.
DR HAMAP; MF_00449; -; 1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBBB48173E788F3 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1005;
Best Local Similarity 24.7%; Pred. No. 7.3;
Matches 43; Conservative 34; Mismatches 62; Indels 35; Gaps 7;

```
Qy           3 DID-IDNVLN--------------LEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQR 47
               ||| :| :||               | ||   |  :  :|:  |    | |||     :::
Db         342 DIDNLDTLLNKIKDEIERVETIKDLLEELKNLNEEIEKIEKYKRICEECKEY-----YEK 396

Qy          48 FL-IIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI----------SITNGDKE 96
               :| :      |  |  | :|       |:  ::|:||  : ::          || |  ||
Db         397 YLELEEKAVEYNKLTLEYITLLQEKKSIEKNINDLETRINKLLEETKNIDIESIENSLKE 456

Qy          97 VEDYEK---NIKKARNKL-RVIASITKETWKIDSLDNLVKEVGGTLQVSENPDD 146
               :|: :|   |::| : :| : :  |  |  ::  : : :||| |   : : | |
Db         457 IEEKKKVLENLQKEKIELNKKLGEINSEIKRLKKILDELKEVEGKCPLCKTPID 510
```

RESULT 5
Y373_HUMAN
ID Y373_HUMAN STANDARD; PRT; 1539 AA.
AC O15078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AB002371; BAA20828.1; -.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 18 1514 COILED COIL (POTENTIAL).
SQ SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;

Query Match 11.3%; Score 87; DB 1; Length 1539;
Best Local Similarity 20.8%; Pred. No. 27;
Matches 31; Conservative 36; Mismatches 64; Indels 18; Gaps 4;

```
Qy          12 LEEEQYELGFK--EGQ---------IQGTKDQYLEGKEYGYQTGFQRFLI------IGYI 54
               |:|::  | :  ||:          ||  : |:         |  | : :|      :  :
Db         302 LDEKEQALYYARLEGRNRAKHLRQTIQSLRRQFSGALPLAQQEKFSKTMIQLQNDKLKIM 361

Qy          55 QELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKAR-NKLRV 113
               ||:      | :  | :  :   |  ||:::: :  | | ::| ::   |:: |  :|::
Db         362 QEMKNSQQEHRNMENKTLEMELKLKGLEELISTLKDTKGAQKVINWHMKIEELRLQELKL 421

Qy         114 IASITKETWKIDSLDNLVKEVGGTLQVSE 142
                 :|:  :|  |:|:: |   |:  |
Db         422 NRELVKDKEEIKYLNNIISEYERTISSLE 450
```

RESULT 6
RA50_THEMA

```
ID   RA50_THEMA     STANDARD;      PRT;   852 AA.
AC   Q9X1X1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR TM1636.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSBB / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001806; AAD36703.1; -.
DR   PIR; D72230; D72230.
DR   TIGR; TM1636; -.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02463; SMC_N; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      30     37       ATP (BY SIMILARITY).
FT   DOMAIN      155    711       COILED COIL (POTENTIAL).
SQ   SEQUENCE   852 AA;  100001 MW;  31BA9F72A4EC5CD2 CRC64;

  Query Match             10.9%;  Score 84.5;  DB 1;  Length 852;
  Best Local Similarity   21.2%;  Pred. No. 22;
  Matches   32;  Conservative   35;  Mismatches   49;  Indels   35;  Gaps    6;

Qy         11 NLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQ-YN 69
              |::| : :|  |  ::: | :  | |:              :  |::  |  |::
Db        452 NIDEFE-KLDQKRSELENTLNVLKERKK-------------SLSSLIEDLLMKIEEGKK 496

Qy         70 NSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVI--------ASITKET 121
              |  |:|| :  :|: : ::  :      || |: : : | ||| |           ||
Db        497 NLKSIRNQIEKIEEELHRLGYS------EDLEEKLDEKRKKLRKIEEERHSISQKITAAD 550

Qy        122 WKIDSLDNLVKEVGGTLQ-----VSENPDDM 147
               :|  ::| :||: | ::     : |  ::|
Db        551 VQISQIENQLKEIKGEIEAKRETLKEQREEM 581


RESULT 7
RA50_PYRHO
ID   RA50_PYRHO     STANDARD;      PRT;   879 AA.
AC   O58687;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR PH0929.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000004; BAA30025.1; -.
DR   PIR; C71083; C71083.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02483; SMC_C; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   SMART; SM00382; AAA; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      30     37       ATP (BY SIMILARITY).
FT   DOMAIN      141    744       COILED COIL (POTENTIAL).
SQ   SEQUENCE   879 AA;  103673 MW;  FD4E30FD1BBCDB29 CRC64;

  Query Match             10.8%;  Score 83.5;  DB 1;  Length 879;
  Best Local Similarity   19.8%;  Pred. No. 27;
  Matches   37;  Conservative   40;  Mismatches   61;  Indels   49;  Gaps    6;

Qy          1 MSDIDIDNVLNLEEEQYELGFKE-----GQIQGTKDQYLEGKEYGYQT------------ 43
              :| |:::  |  ::|:|||   |     |:::  |  :  |  :|  ::
Db        512 LSKINLED-LKRDKEEYELLKSESNKLKGEVESLKKEVNELNDYKNESTKLEIEIDKAKK 570

Qy         44 ------------GFQRF-LIIGYIQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQI-- 88
                         ||:    : | |:|| ||   :|:  |     ||: | :|:|   ::
Db        571 ELSEIEDRLLRLGFKTIDELSGRIRELEKFHNKYIEAKNAEKELRDILESLKDEREELDK 630

Qy         89 -------------SITNGDKEVE---DYEKNIKKARNKLRVIASITKETWKIDSLDNLVK 132
                          :|:   |::   | :|  :|    :::    |      |:: |:
Db        631 AFEELAKIETDIEKVTSQLNELQRKFDQKKYEEKREKMMKLSMEIKGLETKLEELERRRD 690

Qy        133 EVGGTLQ 139
               |:  |::
Db        691 EIKSTIE 697
```

```
RESULT 8
RA50_PYRAB
ID   RA50_PYRAB     STANDARD;      PRT;   880 AA.
AC   Q9UZC8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   DNA double-strand break repair rad50 ATPase.
GN   RAD50 OR PYRAB12200 OR PAB0812.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC       rad50/mre11 complex possesses single-strand endonuclease activity
CC       and ATP-dependent double-strand-specific exonuclease activity.
CC       Rad50 provides an ATP-dependent control of mre11 by unwinding
CC       and/or repositioning DNA ends into the mre11 active site (By
CC       similarity).
CC   -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ248286; CAB50131.1; -.
DR   PIR; F75103; F75103.
DR   HAMAP; MF_00449; -; 1.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF04423; Rad50_zn_hook; 1.
DR   Pfam; PF02463; SMC_N; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      30     37       ATP (BY SIMILARITY).
FT   DOMAIN      144    745       COILED COIL (POTENTIAL).
SQ   SEQUENCE   880 AA;  103970 MW;  FDB177EC7E026479 CRC64;

  Query Match             10.8%;  Score 83.5;  DB 1;  Length 880;
  Best Local Similarity   20.1%;  Pred. No. 27;
  Matches   38;  Conservative   35;  Mismatches   49;  Indels   67;  Gaps    8;

Qy         11 NLEEEQYELGFKEGQIQGTKDQY--LEGKEYGYQTGFQRF--------LI---------- 50
              |||    ||  || : :|  :::  |:|:  | :   :|         ||
Db        517 NLE----ELEQKEREFEGLNEEFNKLKGELLGLERDLKRIKALEGRRKLIEEKVRKAKEE 572

Qy         51 ----------IGY---------IQELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQIS-- 89
                       :|:         |||| :|   :::   : | ||   | ||    ::
Db        573 LENLHRQLRELGFESVEELNLRIQELEEFHDKYVEAKKSESELRELKNKLEKEKTELDQA 632

Qy         90 ---ITNGDKEVEDYEKNIKKARNKL----------------RVIASITKETWKIDSLDNL 130
                 : : : |:|: |  :|   :|                | ::|:|    ::: |
Db        633 FEMLADVENEIEEKEAKLKDLESKFNEEEYEEKRERLVKLEREVSSLTA---RLEELKKS 689

Qy        131 VKEVGGTLQ 139
                |:::  ||:
Db        690 VEQIKATLR 698


RESULT 9
TRP6_MOUSE
ID   TRP6_MOUSE     STANDARD;      PRT;   930 AA.
AC   Q61143; Q9Z2J1;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Short transient receptor potential channel 6 (TrpC6) (Calcium entry
DE   channel).
GN   TRPC6 OR TRRP6 OR TRP6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98037793; PubMed=9368034;
RA   Boulay G., Zhu X., Peyton M., Jiang M., Hurst R., Stefani E.,
RA   Birnbaumer L.;
RT   "Cloning and expression of a novel mammalian homolog of Drosophila
RT   transient receptor potential (Trp) involved in calcium entry secondary
RT   to activation of receptors coupled by the Gq class of G protein.";
RL   J. Biol. Chem. 272:29672-29680(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2;
RX   MEDLINE=99158172; PubMed=10050885;
RA   Buess M., Engler O., Hirsch H.H., Moroni C.;
RT   "Search for oncogenic regulators in an autocrine tumor model using
RT   differential display PCR: identification of novel candidate genes
RT   including the calcium channel mtrp6.";
RL   Oncogene 18:1487-1494(1999).
RN   [3]
RP   SEQUENCE OF 631-739 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96234226; PubMed=8646775;
RA   Zhu X., Jiang M., Peyton M., Boulay G., Hurst R., Stefani E.,
RA   Birnbaumer L.;
RT   "trp, a novel mammalian gene family essential for agonist-activated
RT   capacitative Ca2+ entry.";
RL   Cell 85:661-671(1996).
CC   -!- FUNCTION: THOUGHT TO FORM A RECEPTOR-ACTIVATED NON-SELECTIVE
CC       CALCIUM PERMEANT CATION CHANNEL. PROBABLY IS OPERATED BY A
CC       PHOSPHATIDYLINOSITOL SECOND MESSENGER SYSTEM ACTIVATED BY RECEPTOR
CC       TYROSINE KINASES OR G-PROTEIN COUPLED RECEPTORS. ACTIVATED BY
CC       DIACYLGLYCEROL (DAG) IN A MEMBRANE-DELIMITED FASHION,
CC       INDEPENDENTLY OF PROTEIN KINASE C. SEEMS NOT TO BE ACTIVATED BY
CC       INTRACELLULAR CALCIUM STORE DEPLETION.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- TISSUE SPECIFICITY: LUNG AND BRAIN.
CC   -!- PTM: N-GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO THE TRANSIENT RECEPTOR FAMILY. STRPC
CC       SUBFAMILY.
CC   -!- SIMILARITY: Contains 3 ANK repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U49069; AAC06146.1; -.
DR   EMBL; AF057748; AAC64394.1; -.
DR   MGD; MGI:109523; Trpc6.
DR   GO; GO:0015279; F:store-operated calcium channel activity; IDA.
```

```
DR   GO; GO:0007204; P:cytosolic calcium ion concentration elevation; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR002153; Trans_receptor.
DR   InterPro; IPR004729; Trp_CaChannel.
DR   InterPro; IPR005462; TRPChannel6.
DR   Pfam; PF00023; ank; 4.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR01097; TRNSRECEPTRP.
DR   PRINTS; PR01647; TRPCHANNEL6.
DR   SMART; SM00248; ANK; 2.
DR   TIGRFAMs; TIGR00870; trp; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 1.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   Ionic channel; Transmembrane; Ion transport; Calcium channel;
KW   ANK repeat; Repeat; Glycoprotein.
FT   DOMAIN        1    437       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    438    458       POTENTIAL.
FT   DOMAIN      459    486       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    487    507       POTENTIAL.
FT   DOMAIN      508    520       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    521    541       POTENTIAL.
FT   DOMAIN      542    591       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    592    612       POTENTIAL.
FT   DOMAIN      613    635       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    636    656       POTENTIAL.
FT   DOMAIN      657    705       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    706    726       POTENTIAL.
FT   DOMAIN      727    930       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      131    160       ANK 1.
FT   REPEAT      162    188       ANK 2.
FT   REPEAT      217    246       ANK 3.
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      3     56       MISSING (IN REF. 2).
FT   CONFLICT    105    105       V -> A (IN REF. 2).
FT   CONFLICT    114    114       W -> R (IN REF. 2).
FT   CONFLICT    134    134       N -> D (IN REF. 2).
FT   CONFLICT    184    184       S -> A (IN REF. 2).
FT   CONFLICT    371    371       D -> Y (IN REF. 2).
FT   CONFLICT    436    437       PR -> RG (IN REF. 2).
FT   CONFLICT    905    905       S -> T (IN REF. 2).
SQ   SEQUENCE   930 AA;  106732 MW;  CF21A426972732F3 CRC64;

  Query Match             10.8%;  Score 83.5;  DB 1;  Length 930;
  Best Local Similarity   25.3%;  Pred. No. 29;
  Matches   37;  Conservative   29;  Mismatches   49;  Indels   31;  Gaps    6;

Qy         19 LGFKEGQ---IQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF-----WLSHIDQYNN 70
              | ||:     ||| |  :  |   :   ::| | |  ::| ||      |:|  | :
Db        777 LKFKKWMCELIQGQKQGFQEDAEMNKRNEEKKFGISGSHEDLSKFSLDKNQLAHNKQSST 836

Qy         71 SSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKARNKLRVIASITKETWKIDSLDNL 130
              ||   |||         | :|  ::   |:| :|:   :  : | | ||: :::  :
Db        837 RSSEDYHLN---------SFSNPPRQ---YQKIMKRLIKRYVLQAQIDKESDEVN--EGE 882

Qy        131 VKEVGGTL---------QVSENPDDM 147
              :||:   :         : |:| :|:
Db        883 LKEIKQDISSLRYELLEEKSQNSEDL 908


RESULT 10
SMC1_YEAST
ID   SMC1_YEAST     STANDARD;      PRT;  1225 AA.
AC   P32908;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Structural maintenance of chromosome 1 (DA-box protein SMC1).
GN   SMC1 OR CHL10 OR YFL008W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTANTS SMC1-1 AND SMC1-2.
RX   MEDLINE=94103320; PubMed=8276886;
RA   Strunnikov A.V., Larionov V.L., Koshland D.;
RT   "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT   protein is required for nuclear division and defines a new ubiquitous
RT   protein family.";
RL   J. Cell Biol. 123:1635-1648(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
RN   [3]
RP   IDENTIFICATION IN A COHESIN COMPLEX WITH SMC3; IRR1 AND MCD1, AND
RP   INTERACTION OF THE COHESIN COMPLEX WITH SCC2.
RX   PubMed=9990856;
RA   Toth A., Ciosk R., Uhlmann F., Galova M., Schleiffer A., Nasmyth K.;
RT   "Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to
RT   establish cohesion between sister chromatids during DNA replication.";
RL   Genes Dev. 13:320-333(1999).
RN   [4]
RP   IDENTIFICATION IN A COHESIN COMPLEX WITH SMC3; MCD1 AND IRR1, AND
RP   STRUCTURE.
RX   PubMed=11983169;
RA   Haering C.H., Loewe J., Hochwagen A., Nasmyth K.;
RT   "Molecular architecture of SMC proteins and the yeast cohesin
RT   complex.";
RL   Mol. Cell 9:773-788(2002).
CC   -!- FUNCTION: Involved in chromosome cohesion during cell cycle and in
CC       DNA repair. Central component of cohesin complex. The cohesin
CC       complex is required for the cohesion of sister chromatids after
CC       DNA replication. The cohesin complex apparently forms a large
CC       proteinaceous ring within which sister chromatids can be trapped.
CC       At anaphase, the complex is cleaved and dissociates from
CC       chromatin, allowing sister chromatids to segregate.
CC   -!- SUBUNIT: Cohesin complexes are composed of the SMC1 and SMC3
CC       heterodimer attached via their hinge domain, MCD1/SCC1 which link
CC       them, and IRR1/SCC3, which interacts with MCD1. The cohesin
CC       complex also interacts with SCC2, which is required for its
CC       association with chromosomes.
CC   -!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
CC       Before prophase it is scattered along chromosome arms. At
CC       anaphase, the MCD1 subunit of the cohesin complex is cleaved,
CC       leading to the dissociation of the complex from chromosomes,
CC       allowing chromosome separation.
CC   -!- DOMAIN: The flexible hinge domain, which separates the large
CC       intramolecular coiled coil regions, allows the heterotypic
CC       interaction with the corresponding domain of SMC3, forming a V-
CC       shaped heterodimer. The two heads of the heterodimer are then
CC       connected by different ends of the cleavable MCD1 protein, forming
CC       a ring structure.
CC   -!- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00602; AAA16595.1; -.
DR   EMBL; D50617; BAA09230.1; -.
DR   PIR; A49464; A49464.
DR   SGD; S0001886; SMC1.
DR   GO; GO:0008278; C:cohesin complex; IDA.
```

```
DR   GO; GO:0003680; F:AT DNA binding activity; IDA.
DR   GO; GO:0000217; F:DNA secondary structure binding activity; IDA.
DR   GO; GO:0003690; F:double-stranded DNA binding activity; IDA.
DR   GO; GO:0000070; P:mitotic chromosome segregation; IMP.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02463; SMC_N; 1.
DR   Pfam; PF02483; SMC_C; 1.
KW   Mitosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   NP_BIND      33     40       ATP (POTENTIAL).
FT   DOMAIN      173    489       COILED COIL (POTENTIAL).
FT   DOMAIN      490    678       FLEXIBLE HINGE.
FT   DOMAIN      679   1063       COILED COIL (POTENTIAL).
FT   DOMAIN     1057   1061       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1129   1164       ALA/ASP-RICH (DA-BOX).
FT   MUTAGEN     173    173       S->L: IN TS MUTANT SMC1-2.
FT   MUTAGEN     458    458       N->D: IN TS MUTANT SMC1-1.
SQ   SEQUENCE   1225 AA;  141279 MW;  B504017AA0ECCA8C CRC64;

  Query Match           10.7%;  Score 82.5;  DB 1;  Length 1225;
  Best Local Similarity  18.3%;  Pred. No. 48;
  Matches   28;  Conservative   36;  Mismatches   50;  Indels   39;  Gaps    5;

Qy          5 DIDN----VLNLEEEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKF 60
              |::|    : :|||::| :  | | |:   :::                  :: || |
Db        852 DLENAQVEMKSLEEQEYAIEMKIGSIESKLEEHK----------------NHLDELQKK 894

Qy         61 WLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKE--VEDYE----------KNIKKAR 108
              :::   : |:|  :       |||: : : :   :::   || |          || | :
Db        895 FVTKQSELNSSEDI------LEDMNSNLQVLKRERDGIKEDIEKFDLERVTALKNCKISN 948

Qy        109 NKLRVIASITKETWKIDSLDNLVKEVGGTLQVS 141
                : : :  | :   | | ||    :  :: ::
Db        949 INIPISSETTIDDLPISSTDNEAITISNSIDIN 981


RESULT 11
MYSB_CAEEL
ID   MYSB_CAEEL     STANDARD;      PRT;  1966 AA.
AC   P02566;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain B (MHC B).
GN   UNC-54 OR MYO-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83273600; PubMed=6576334;
RA   Karn J., Brenner S., Barnett L.;
RT   "Protein structural domains in the Caenorhabditis elegans unc-54
RT   myosin heavy chain gene are not separated by introns.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN   [2]
RP   SEQUENCE OF 850-1966 FROM N.A.
RX   MEDLINE=82272395; PubMed=7202124;
RA   McLachlan A.D., Karn J.;
RT   "Periodic charge distributions in the myosin rod amino acid sequence
RT   match cross-bridge spacings in muscle.";
RL   Nature 299:226-231(1982).
RN   [3]
RP   SEQUENCE OF 1876-1966 FROM N.A.
RX   MEDLINE=83232892; PubMed=6571695;
RA   Wills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,
RA   Waterston R.H.;
RT   "The genes sup-7 X and sup-5 III of C. elegans suppress amber
RT   nonsense mutations via altered transfer RNA.";
RL   Cell 33:575-583(1983).
CC   -!- FUNCTION: MUSCLE CONTRACTION.
CC   -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC       HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC       AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC       CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC       CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC   -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC       ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC   -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC       MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC       SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC       SUBFRAGMENT (S2).
CC   -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC       C.ELEGANS.
CC   -!- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
CC       WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J01050; AAA28124.1; -.
DR   EMBL; V01494; CAA24738.1; -.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00242; MYSc; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Methylation; Alkylation; Multigene family.
FT   DOMAIN        1    850       MYOSIN HEAD-LIKE.
FT   DOMAIN      851   1966       COILED COIL (POTENTIAL).
FT   DOMAIN      851   1164       ALPHA-HELICAL TAILPIECE (S2).
FT   DOMAIN     1165   1176       HINGE.
FT   DOMAIN     1165   1966       LIGHT MEROMYOSIN (LMM).
FT   NP_BIND     177    184       ATP (BY SIMILARITY).
FT   DOMAIN      665    687       ACTIN-BINDING.
FT   DOMAIN      769    783       ACTIN-BINDING.
FT   MOD_RES     128    128       METHYLATION (TRI-) (POTENTIAL).
FT   MOD_RES     705    705       ALKYLATION (SH-1).
FT   MOD_RES     715    715       ALKYLATION (SH-2).
FT   CONFLICT   1337   1337       E -> R (IN REF. 2).
FT   CONFLICT   1880   1880       I -> L (IN REF. 2).
SQ   SEQUENCE   1966 AA;  225125 MW;  B66F0BB2FE27B67F CRC64;

  Query Match           10.6%;  Score 82;  DB 1;  Length 1966;
  Best Local Similarity  20.9%;  Pred. No. 89;
  Matches   37;  Conservative   35;  Mismatches   59;  Indels   46;  Gaps    6;

Qy          7 DNVLNLEEEQYELGFKEGQIQGTKDQYLE--------GKEYGYQTGFQRFLIIGYIQELM 58
              |  ::| : : |    |: ||:  :|:  :           ||  :|    |         :||
Db        968 DLEMSLRKAESEKQSKDHQIRSLQDEMQQQDEAIAKLNKEKKHQEEINR--------KLM 1019

Qy         59 KFWLSHIDQYNN----SSSLRNHLNNLEDIM----------------------AQISIT 91
              :   |  |: |:    : |    |:|||  :                      || :|
Db       1020 EDLQSEEDKGNHQNKVKAKLEQTLDDLEDSLEREKRARADLDKQKRKVEGELKIAQENID 1079

Qy         92 NGDKEVEDYEKNIKKARNKLRVIAS-ITKETWKIDSLDNLVKEVGGTLQVSENPDDM 147
                ::  | | |:||  ::|  ::| :  |   :  |    :|:  |  ::||  :::
Db       1080 ESGRQRHDLENNLKKKESELHSVSSRLEDEQALVSKLQRQIKD--GQSRISELEEEL 1134
```

```
RESULT 12
K6PF_SALTY
ID   K6PF_SALTY     STANDARD;      PRT;   320 AA.
AC   Q8XGI9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
DE   (Phosphohexokinase).
GN   PFKA OR STM4062 OR STY3809 OR T3557.
OS   Salmonella typhimurium, and
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602, 601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC       fructose 1,6-bisphosphate.
CC   -!- PATHWAY: Key control step of glycolysis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the phosphofructokinase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE008889; AAL22902.1; -.
DR   EMBL; AL627279; CAD09562.1; -.
DR   EMBL; AE016846; AAO71062.1; -.
DR   StyGene; SG?????; pfkA.
DR   HAMAP; MF_00339; -; 1.
DR   InterPro; IPR000023; Ppfruckinase.
DR   Pfam; PF00365; PFK; 1.
DR   PRINTS; PR00476; PHFRCTKINASE.
DR   ProDom; PD000707; Ppfruckinase; 1.
DR   PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
KW   Kinase; Transferase; Glycolysis; Complete proteome.
SQ   SEQUENCE   320 AA;  34915 MW;  68B0DDFCF689F420 CRC64;

  Query Match             10.5%;  Score 81;  DB 1;  Length 320;
  Best Local Similarity   25.2%;  Pred. No. 14;
  Matches   35;  Conservative   23;  Mismatches   47;  Indels   34;  Gaps    7;

Qy         18 ELGFKEGQIQGTKDQYLEGKEY--GYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSSLR 75
              |:||    : ||: |  ::| :|  ||       |  :| : |         ||:  ::||
Db        115 EMGFPCIGLPGTIDNDIKGTDYTIGY------FTALGTVVEA-------IDRLRDTSSSH 161

Qy         76 NHLNNLE-------DIMAQISITNGDKEV---------EDYEKNIKK--ARNKLRVIASI 117
                :: :|       |:    :|  | : :          ||    ||   |: |   | :|
Db        162 QRISIVEVMGRYCGDLTLAAAIAGGCEFIVVPEVEFNREDLVAEIKAGIAKGKKHAIVAI 221

Qy        118 TKETWKIDSLDNLV-KEVG 135
              |:    :| | : || |
Db        222 TEHMCDVDELAHFIEKETG 240


RESULT 13
SYK_BORBU
ID   SYK_BORBU      STANDARD;      PRT;   521 AA.
AC   O51603;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LysRS).
GN   LYSS OR BB0659.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=98070760; PubMed=9405621;
RA   Ibba M., Bono J.L., Rosa P.A., Soell D.;
RT   "Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete
RT   Borrelia burgdorferi.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14383-14388(1997).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=20570460; PubMed=11121028;
RA   Soell D., Becker H.D., Plateau P., Blanquet S., Ibba M.;
RT   "Context-dependent anticodon recognition by class I lysyl-tRNA
RT   synthetases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:14224-14228(2000).
CC   -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC       + L-lysyl-tRNA(Lys).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MISCELLANEOUS: Able to charge E.coli tRNA(Lys) in vitro.
CC   -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; AE001167; AAC67006.1; -.
DR   PIR; B70182; B70182.
DR   TIGR; BB0659; -.
DR   HAMAP; MF_00177; -; 1.
DR   InterPro; IPR002904; Lys_tRNA-synt_1c.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF01921; tRNA-synt_1f; 1.
DR   TIGRFAMs; TIGR00467; lysS_arch; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         32     40       "HIGH" REGION.
FT   SITE        280    284       "KMSKS" REGION.
SQ   SEQUENCE   521 AA;  60938 MW;  6AF5A461AED3251D CRC64;

  Query Match             10.5%;  Score 81;  DB 1;  Length 521;
  Best Local Similarity   21.5%;  Pred. No. 24;
  Matches   32;  Conservative   22;  Mismatches   49;  Indels   46;  Gaps    4;

Qy         14 EEQYELGFKEGQIQGTKDQYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQYNNSSS 73
              :|: :  ||  :|      |: |   || ||:  :| | |             || :
Db        345 KEEKKRAFK--RIYELSQPYMPSKRIPYQVGFRHLSVISQIFE-----------NNINK 390

Qy         74 LRNHLNNLEDIMA--------------------QISITNGDKEVEDYEKNIKKARNKL 111
              : |:| |:::                     : |: :   :|  |:| ||| |:|
Db        391 ILNYLKNVQEDQKDKLINKINCAINWIRDFAPEDFKFSLRSKFDNMEILEENSKKAINEL 450

Qy        112 ----------RVIASITKETWKIDSLDNL 130
                       |  | :||   :|:
Db        451 LDFLKKNFEVATEQDIQNEIYKISRENNI 479


RESULT 14
M12_STRPY
ID   M12_STRPY      STANDARD;      PRT;   564 AA.
AC   P19401;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   M protein, serotype 12 precursor (Fragment).
GN   EMM12.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CS24 / Serotype M12;
RX   MEDLINE=88058777; PubMed=2445730;
RA   Robbins J.C., Spanier J.G., Jones S.J., Simpson W.J., Cleary P.P.;
RT   "Streptococcus pyogenes type 12 M protein gene regulation by upstream
RT   sequences.";
RL   J. Bacteriol. 169:5633-5640(1987).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC       OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC       THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC       PHAGOCYTOSIS.
CC   -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC       an amide bond (Potential).
CC   -!- SIMILARITY: TO OTHER M PROTEINS.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; M18269; AAA88573.1; -.
DR   PIR; A40174; A60115.
DR   InterPro; IPR005877; Gpos_YSIRK.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR006192; LPXTG.
DR   InterPro; IPR003345; M_repeat.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02370; M; 9.
DR   Pfam; PF04650; YSIRK_signal; 1.
DR   PRINTS; PR00015; GPOSANCHOR.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
KW   Antigen; Coiled coil; Signal.
FT   SIGNAL       1     41
FT   CHAIN       42    545       M PROTEIN, SEROTYPE 12.
FT   PROPEP     546   >564       REMOVED BY SORTASE (POTENTIAL).
FT   DOMAIN      44    505       COILED COIL (POTENTIAL).
FT   DOMAIN     505    541       GLY/PRO-RICH (CELL WALL-SPANNING).
FT   SITE       542    546       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES    545    545       AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT   NON_TER    564    564
SQ   SEQUENCE   564 AA;  62904 MW;  5F1549DACAA77B46 CRC64;

  Query Match             10.5%;  Score 81;  DB 1;  Length 564;
  Best Local Similarity   22.6%;  Pred. No. 26;
  Matches   35;  Conservative   30;  Mismatches   48;  Indels   42;  Gaps    7;

Qy         13 EEEQYE-LGFKEGQIQGTKD----QYLEGKEYGYQTGFQRFLIIGYIQELMKFWLSHIDQ 67
              |::: | || |  :::   :    || : |  ||: :        |:|:| |    ::|
Db         49 EKQRLEDLGQKFERLKQRSELYLQQYYDNKSNGYKGDW-------YVQQL-KMLNRDLEQ 100

Qy         68 YNNSSSLRNHLNNLEDIMAQISITNGD----------------------KEVEDYEKNI 104
              |  |  |    :| : :: | | |                        ||:|: ||:|
Db        101 AYNELSGEAH----KDALGKLGIDNADLKAKITELEKSVEEKNDVLSQIKKELEEAEKDI 156

Qy        105 KKARNKLRVIASITKETWKIDSLDNLVKEVGGTLQ 139
              :  |      | : :   :|   :|:: :: | ||
Db        157 QFGREV--HAADLLRHKQEIAEKENVISKLNGELQ 189


RESULT 15
Y373_BOVIN
ID   Y373_BOVIN     STANDARD;      PRT;  1453 AA.
AC   Q9TU23;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein KIAA0373.
GN   KIAA0373.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Trachea;
RA   Jovov B., Ripoll P.J., Benos D.J.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; AF176816; AAF00990.1; -.
KW   Hypothetical protein; Coiled coil.
FT   DOMAIN       37   1426       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1453 AA;  169934 MW;  03CBA02A64CF4139 CRC64;
```

Query Match 10.5%; Score 81; DB 1; Length 1453;
Best Local Similarity 20.1%; Pred. No. 76;
Matches 30; Conservative 34; Mismatches 67; Indels 18; Gaps 4;

```
Qy          12 LEEEQYELGFK--EGQ---------IQGTKDQYLEGKEYGYQTGFQRFLI------IGYI 54
               |:|::  |  :   ||:          ||  : |:         |  | : :|      :  :
Db         321 LDEKEQALFYARLEGRNRAKHLRQTIQSLRRQFSGALPLAQQEKFSKTMIQLQNDKLKIM 380

Qy          55 QELMKFWLSHIDQYNNSSSLRNHLNNLEDIMAQISITNGDKEVEDYEKNIKKAR-NKLRV 113
               :|:       |    | :  :   |  |||::: :    | ::|  :   |:: |  :|::
Db         381 EEMKNSQQEHRSLKNKTLEMELKLKGLEDLISTLKDARGAQKVISWHTKIEELRLQELKL 440

Qy         114 IASITKETWKIDSLDNLVKEVGGTLQVSE 142
                 : |:  :|  |:|:: |   |:   |
Db         441 NRELVKDKEEIKYLNNIISEYENTISSLE 469
```

Search completed: February 16, 2004, 14:12:32
Job time : 13 secs